第四章 独立随机变量之和与独立随机变量序列



两个或多个试验的**独立性**的概念, 一定意义上在概率中占中心位置 …… 历史上试验和随机变量的独立性, 曾经是赋予概率论以印迹一种数学概念.

A. H. **柯尔莫戈洛夫,《概率论的基本概念》[32]**

§1. 0–1 律

1. 柯尔莫戈洛夫 0–1 律 熟知, 下面两个级数:

$$\sum_{n=1}^{\infty}\frac{1}{n} \quad 和 \quad \sum_{n=1}^{\infty}(-1)^n\frac{1}{n},$$

其中前一个级数发散, 后一个级数收敛. 现在, 提出如下问题: 对于独立同分布伯努利随机变量序列$\xi_1, \xi_2, \cdots$ 关于级数

$$\sum_{n=1}^{\infty}\frac{\xi_n}{n}, \quad 其中 \quad \mathbf{P}\{\xi_1 = +1\} = \mathbf{P}\{\xi_1 = -1\} = \frac{1}{2},$$

的收敛性有何说法? 换句话说, 假如级数的一般项为 $\pm 1/n$, 其中符号 "+" 和 "−" 与上述序列 $\xi_1, \xi_2, \cdots$ 相对应, 并按随机顺序 "散布", 那么是否可以说 "以 ξ_n/n 为一般项的级数收敛"?

记

$$A_1 = \left\{\omega : \sum_{n=1}^{\infty}\frac{\xi_n(\omega)}{n} \quad 收敛\right\}$$

是使其中的级数收敛 (于某些数值) 的基本事件 ω 的集合, 并且在事先并不知道该集合的概率 $\mathbf{P}(A_1)$ 取何值的情况下, 讨论概率 $\mathbf{P}(A_1)$.

不过, 非常精彩的是, 事先就可以断定: 这一概率只可能取 0 或 1 两个值之一. 这一结果是称做柯尔莫戈洛夫 (A. H. Колмогоров) 0–1 律的推论. 0–1 律的提法和证明是这一节的基本内容.

2. 柯尔莫戈洛夫 0–1 律的证明 设 $(\Omega, \mathscr{F}, \mathbf{P})$ 是概率空间, $\xi_1, \xi_2, \cdots$ 是一随机变量序列. 以 $\mathscr{F}_n^{\infty} = \sigma(\xi_n, \xi_{n+1}, \cdots)$ 表示由随机变量 $\xi_n, \xi_{n+1}, \cdots$ 生成的 σ– 代数, 并且设

$$\mathscr{X} = \bigcap_{n=1}^{\infty}\mathscr{F}_n^{\infty}.$$

由于 $\sigma-$ 代数的交仍然是 $\sigma-$ 代数, 可见 $\mathscr{X}$ 是 $\sigma-$ 代数. 因为对于任意有限数 n, 任何事件 $A \in \mathscr{X}$ 都不依赖于 $\xi_1, \cdots, \xi_n$, 而仅由"序列 $\xi_1, \xi_2, \cdots$ 的无穷远的性质"所决定, 故我们把 $\sigma-$ 代数 $\mathscr{X}$ 称做"尾部的"或"剩余的".

由于对于任何 $k \geqslant 1$, 有

$$A_1 \equiv \left\{\sum_{n=1}^{\infty} \frac{\xi_n}{n} \quad \text{收敛}\right\} = \left\{\sum_{n=k}^{\infty} \frac{\xi_n}{n} \quad \text{收敛}\right\} \in \mathscr{F}_k^{\infty},$$

则 $A_1 = \bigcap\limits_k \mathscr{F}_k^{\infty} \equiv \mathscr{X}$. 同样, 假如 $\xi_1, \xi_2, \cdots$ 是任一随机变量序列, 则

$$A_2 = \left\{\sum_{n=1}^{\infty} \xi_n \quad \text{收敛}\right\} \in \mathscr{X}.$$

下列事件也是"尾部的":

$$A_3 = \{\text{对于无限个 } n, \xi_n \in I_n\} \quad \left(= \varlimsup_n \{\xi_n \in I_n\}\right)$$

其中 $I_n \in \mathscr{B}(\mathbb{R}), n \geqslant 1$;

$$A_4 = \left\{\varlimsup_n \xi_n < \infty\right\};$$

$$A_5 = \left\{\varlimsup_n \frac{\xi_1 + \cdots + \xi_n}{n} < \infty\right\};$$

$$A_6 = \left\{\varlimsup_n \frac{\xi_1 + \cdots + \xi_n}{n} < c\right\};$$

$$A_7 = \left\{\frac{S_n}{n} \quad \text{收敛}\right\};$$

$$A_8 = \left\{\varlimsup_n \frac{S_n}{\sqrt{2n \ln n}} = 1\right\},$$

其中 $S_n = \xi_1 + \cdots + \xi_n$. 另一方面

$$B_1 = \{\text{对于一切} \quad n \geqslant 1, \xi_n = 0\},$$

$$B_2 = \left\{\lim_n (\xi_1 + \cdots + \xi_n) \ \text{存在且小于 } c\right\}$$

都是不属于的 $\mathscr{X}$ 事件的例.

现在假设所考虑的随机变量是独立的. 在此条件下, 由博雷尔 – 坎泰利引理, 可见

$$\mathbf{P}(A_3) = 0 \Leftrightarrow \sum \mathbf{P}\{\xi_n \in I_n\} < \infty,$$

$$\mathbf{P}(A_3) = 1 \Leftrightarrow \sum \mathbf{P}\{\xi_n \in I_n\} = \infty.$$

这样, 以级数 $\sum \mathbf{P}\{\xi_n \in I_n\}$ 是否收敛为转移, 事件 A_3 的概率只有 0 或 1 两个可能值. 这一命题称做博雷尔"0–1"律, 它是如下命题的特殊情形.

定理 1 (柯尔莫戈洛夫 "0–1" 律) 设 $\xi_1, \xi_2, \cdots$ 是独立随机变量序列, 而 $A \in \mathscr{X}$. 那么, 概率 $\mathbf{P}(A)$ 只有 0 和 1 两个可能值.

证明 证明的思路如下: 为证明每一"尾部" 事件 A 与它自己独立, 只需证明 $\mathbf{P}(A \bigcap A) = \mathbf{P}(A) \times \mathbf{P}(A), \mathbf{P}(A) = \mathbf{P}^2(A)$, 故 $\mathbf{P}(A) = 0$ 或 1.

若 $A \in \mathscr{X}$, 则 $A \in \mathscr{F}_1^\infty = \sigma\{\xi_1, \xi_2, \cdots\} = \sigma(\bigcup \mathscr{F}_1^n)$, 其中 $\mathscr{F}_1^n = \sigma(\xi_1, \cdots, \xi_n)$, 并且存在集合 $A_n \in \mathscr{F}_1^n, n \geqslant 1$ (第二章 §3, 练习题 8), 使 $\mathbf{P}(A \triangle A_n) \to 0, n \to \infty$. 由此可见,

$$\mathbf{P}(A_n) \to \mathbf{P}(A), \quad \mathbf{P}(A_n \bigcap A) \to \mathbf{P}(A). \tag{1}$$

但是, 假如 $A \in \mathscr{X}$, 则对于每一个 $n \geqslant 1$, 事件 A_n 和 A独立:

$$\mathbf{P}(A \bigcap A_n) = \mathbf{P}(A)\mathbf{P}(A_n),$$

由于 (1), 由此可见 $\mathbf{P}(A) = \mathbf{P}^2(A)$, 故 $\mathbf{P}(A) = 0$ 或 1. □

系 设 η 是关于 "尾部" $\sigma-$ 代数 $\mathscr{X}$ 可测的随机变量, 即 $\{\eta \in B\} \in \mathscr{X}, B \in \mathscr{B}(\mathbb{R})$, 则 η 是退化随机变量, 即存在常数 c, 使 $\mathbf{P}\{\eta = c\} = 1$.

3. 柯尔莫戈洛夫 0–1 律的应用 下面引进的定理 2 是柯尔莫戈洛夫 "0–1 律" 的非平凡应用.

设 $\xi_1, \xi_2, \cdots$ 是独立伯努利随机变量序列: $\mathbf{P}\{\xi_n = 1\} = p, \mathbf{P}\{\xi_n = -1\} = q, p + q = 1, n \geqslant 1$, 而 $S_n = \xi_1 + \cdots + \xi_n$. 直观上明显, 对于对称 $(p = 1/2)$ 的情形, 随机游动 $S_n (n \geqslant 1)$ 的 "典型" 轨道无限多次地经过 0, 而对于 $p \neq 1/2$ 的情形 "趋向" 无穷. 下面是这一结果的确切表述.

定理 2 a) 如果 $p = 1/2$, 则 $\mathbf{P}\{$对于无限多个 $n, S_n = 0\} = 1$.

b) 如果 $p \neq 1/2$, 则 $\mathbf{P}\{$对于无限多个 $n, S_n = 0\} = 0$.

证明 首先注意到, $B = \{$对于无限多个 $n, S_n = 0\}$ 不是"尾部" 事件, 即 $B \notin \mathscr{X} = \bigcap \mathscr{F}_n^\infty, \mathscr{F}_n^\infty = \sigma\{\xi_n, \xi_{n+1}, \cdots\}$. 因此, 原则上不能说明事件 B 的概率只有 0 或 1 两个可能值.

利用博雷尔 - 坎泰利 (E. Borel – F. P. Cantelli) 引理 (的第一部分), 容易证明命题 b). 事实上, 如果 $B_{2n} = \{S_{2n} = 0\}$, 则可见斯特林 (J. Stirling) 公式 (第一章 §2 第 4 小节), 有

$$\mathbf{P}(B_{2n}) = \mathrm{C}_{2n}^n p^n q^n \sim \frac{(4pq)^n}{\sqrt{n\pi}},$$

从而 $\sum \mathbf{P}(B_{2n}) < \infty$. 因此, $\mathbf{P}\{$对于无限多个 $n, S_n = 0\} = 0$.

由于$A \subseteq B$, 为证明命题 a), 只需证明事件

$$A = \left\{ \overline{\lim} \frac{S_n}{\sqrt{n}} = \infty, \quad \underline{\lim} \frac{S_n}{\sqrt{n}} = -\infty \right\}$$

的概率等于 1.

设 $A_c = A'_c \bigcap A''_c$, 其中

$$A'_c = \left\{\overline{\lim_n}\frac{S_n}{\sqrt{n}} \geqslant c\right\}, \quad A''_c = \left\{\underline{\lim_n}\frac{S_n}{\sqrt{n}} \leqslant -c\right\}.$$

那么, 当 $c \to \infty$ 时 $A_c \downarrow A$; 这时 A 以及所有 A'_c, A''_c 都是尾部事件. 现在验证, 对于每一个 $c, \mathbf{P}(A'_c) = \mathbf{P}(A''_c) = 1$. 由于 $A'_c \in \mathscr{X}, A''_c \in \mathscr{X}$, 故只需证明 $\mathbf{P}(A'_c) > 0, \mathbf{P}(A''_c) > 0$. 然而, 根据练习题 5 和棣莫弗 – 拉普拉斯定理 (A. DéMoivre- P. S. Laplace, 第一章 §6), 有

$$\mathbf{P}\left\{\underline{\lim_n}\frac{S_n}{\sqrt{n}} \leqslant -c\right\} = \mathbf{P}\left\{\overline{\lim_n}\frac{S_n}{\sqrt{n}} \geqslant c\right\} \geqslant \overline{\lim_n}\mathbf{P}\left\{\frac{S_n}{\sqrt{n}} \geqslant c\right\} > 0.$$

于是, 对于一切 $c > 0, \mathbf{P}(A_c) = 1$, 因而 $\mathbf{P}(A) = \lim_{c\to\infty} \mathbf{P}(A_c) = 1$. □

4. 休伊特和塞维治 0–1 律 我们再次强调, 事件 $B = \{$对于无限多个 $n, S_n = 0\}$ 不是 "尾部" 事件. 不过, 由定理 2 可见, 对于伯努利概型, 这一事件的概率同 "尾部" 事件的概率一样, 只有 0 或 1 两个可能值. 原来, 这种情况并不偶然, 它是所谓*休伊特* (E. Hewitt) *和塞维治* (I. R. Sevage) *0–1 律的推论*. 对于独立同分布随机变量, 休伊特和塞维治 0–1 律, 将定理 1 的结果推广到所谓 "可交换" 事件类 (包括 "尾部" 事件类).

现在给出必要的定义. 集合 $(1, 2, \cdots)$ 到自身的单值影射 $(\pi_1, \pi_2, \cdots)$ 称做*有限可交换的*, 如果对于一切 (只可能有有限个除外) $n, \pi_n = n$.

如果 $\xi = (\xi_1, \xi_2, \cdots)$ 是随机变量序列, 则以 $\pi(\xi)$ 表示 $(\xi_{\pi_1}, \xi_{\pi_2}, \cdots)$ 是随机变量序列.

如果事件 $A = \{\xi \in B\}, B \in \mathscr{B}(\mathbb{R}^\infty)$, 则以 $\pi(A)$ 表示事件 $\{\pi(\xi) \in B\}, B \in \mathscr{B}(\mathbb{R}^\infty)$.

称事件 $A = \{\xi \in B\}, B \in \mathscr{B}(\mathbb{R}^\infty)$ 为*可交换的*, 如果对于任意有限排列 π, 事件 $\pi(A)$ 与 A 重合.

事件 $A = \{$对于无限多个 $n, S_n = 0\}$, 其中 $S_n = \xi_1 + \cdots + \xi_n$, 是可交换事件的例. 此外可以证明 (练习题 4), 属于 "尾部" $\sigma-$ 代数 $\mathscr{X}(S) = \bigcap \mathscr{F}_n^\infty(S)$ 的每一个事件是可交换的, 其中 $\mathscr{F}_n^\infty(S) = \sigma\{\omega : S_n, S_{n+1}, \cdots\}$ 是随机变量 $S_1 = \xi_1, S_2 = \xi_1 + \xi_2, \cdots$ 生成的 $\sigma-$ 代数.

定理 3 (休伊特和塞维治 0–1 律) *设 $\xi = (\xi_1, \xi_2, \cdots)$ 是独立同分布随机变量序列, 而 $A = \{\xi \in B\}$ 是可交换事件, 则 $\mathbf{P}(A) = 0$ 或 1.*

证明 设 $A = \{\xi \in B\}$ 是可交换事件. 选择一集合 $B_n \in \mathscr{B}(\mathbb{R}^n)$, 使 (见第二章 §3 练习题 8)

$$\mathbf{P}(A \triangle A_n) \to 0, \quad n \to \infty, \tag{2}$$

其中 $A_n = \{\omega : (\xi_1, \cdots, \xi_n) \in B_n\}$.

由于随机变量 $\xi_1, \xi_2, \cdots$ 独立且同分布, 可见概率分布 $P_\xi(B) \equiv \mathbf{P}\{\xi \in B\}$ 和 $P_{\pi_n(\xi)}(B) = \mathbf{P}\{\pi_n(\xi) \in B\}$, 其中 $\pi_n(\xi) = (\xi_{n+1}, \cdots, \xi_{2n}, \xi_1, \cdots, \xi_n, \xi_{2n+1}, \xi_{2n+2}, \cdots)$, 对于任何 $n \geqslant 1$ 相同. 因此,

$$\mathbf{P}(A \triangle A_n) = P_\xi(B \triangle B_n) = P_{\pi_n(\xi)}(B \triangle B_n). \tag{3}$$

因为 $A = \{\xi \in B\}$ 是可交换事件, 故

$$A \equiv \{\xi \in B\} \equiv \pi_n(A) \equiv \{\pi_n(\xi) \in B\}.$$

从而

$$\begin{aligned} P_{\pi_n(\xi)}(B \triangle B_n) &= \mathbf{P}(\{\pi_n(\xi) \in B\} \triangle \{\pi_n(\xi) \in B_n\}) \\ &= \mathbf{P}(\{\xi \in B\} \triangle \{\pi_n(\xi) \in B_n\}) = \mathbf{P}(A \triangle \pi_n(A_n)). \end{aligned} \tag{4}$$

这样, 由式 (3) 和 (4), 有

$$\mathbf{P}(A \triangle A_n) = \mathbf{P}(A \triangle \pi_n(A_n)). \tag{5}$$

因为 (2) 式, 由此可见

$$\mathbf{P}(A \triangle \{A_n \bigcap \pi_n(A_n)\}) \to 0, \quad n \to \infty. \tag{6}$$

因而由式 (2), (5) 和 (6), 可得当 $n \to \infty$ 时, 有

$$\begin{gathered} \mathbf{P}(A_n) \to \mathbf{P}(A), \quad \mathbf{P}(\pi_n(A_n)) \to \mathbf{P}(A), \\ \mathbf{P}\{A_n \bigcap \pi_n(A_n)\} \to \mathbf{P}(A). \end{gathered} \tag{7}$$

其次, 由于随机变量 $\xi_1, \xi_2, \cdots$ 的独立性, 有

$$\begin{aligned} \mathbf{P}\{A_n \bigcap \pi_n(A_n)\} &= \mathbf{P}\{(\xi_1, \cdots, \xi_n) \in B_n, (\xi_{n+1}, \cdots, \xi_{2n}) \in B_n\} \\ &= \mathbf{P}\{(\xi_1, \cdots, \xi_n) \in B_n\}\mathbf{P}\{(\xi_{n+1}, \cdots, \xi_{2n}) \in B_n\} = \mathbf{P}(A_n)\mathbf{P}(\pi_n(A_n)). \end{aligned}$$

根据性质 (7), 由此可见

$$\mathbf{P}(A) = \mathbf{P}^2(A),$$

于是 $\mathbf{P}(A) = 0$ 或 1. □

5. 练习题

1. 证明定理 1 的系.

2. 设 $(\xi_n)_{n \geqslant 1}$ 是独立随机变量序列, 证明 $\overline{\lim}\xi_n$ 和 $\underline{\lim}\xi_n$ 是退化随机变量.

3. 设 $(\xi_n)_{n \geqslant 1}$ 是独立随机变量序列, $S_n = \xi_1 + \cdots + \xi_n$, 而 b_n 是常数且 $0 < b_n \uparrow \infty$. 证明

$$\overline{\lim_n} \frac{S_n}{b_n} \quad \text{和} \quad \underline{\lim_n} \frac{S_n}{b_n}$$

是退化随机变量.

4. 设 $S_n = \xi_1 + \cdots + \xi_n, n \geqslant 1$, 而, $\mathscr{X}(S) = \bigcap \mathscr{F}_n^\infty(S), \mathscr{F}_n^\infty(S) = \sigma\{\omega : S_n, S_{n+1}, \cdots\}$, 证明 $\mathscr{X}(S)$ 中的每一个事件都是可交换的.

5. 设 $(\xi_n)_{n\geqslant 1}$ 是独立随机变量序列, 证明对于任意常数 c, 有 $\{\overline{\lim}\xi_n \geqslant c\} \supseteq \{\underline{\lim}\xi_n \geqslant c\}$.

6. 举一概率严格大于 0 且严格小于 1 的《尾部》事件的例.

7. 设 $\xi_1, \xi_2, \cdots$ 是独立随机变量, $\mathbf{E}\xi_n = 0, \mathbf{E}\xi_n^2 = 1, n \geqslant 1$, 且服从中心极限定理 ($\mathbf{P}\{S_n/\sqrt{n} \leqslant x\} \to \Phi(x), x \in \mathbb{R}$, 其中 $S_n = \xi_1 + \cdots + \xi_n$, 而 $\Phi(x)$ 是标准正态分布函数), 证明

$$\overline{\lim_{n\to\infty}}\, n^{1/2} S_n = +\infty \quad (\mathbf{P} - \text{a.c.}).$$

特别, 对于独立同分布随机变量序列 (若 $\mathbf{E}\xi_1 = 0, \mathbf{E}\xi_1^2 = 1$), 这一性质仍然成立.

8. 设 $\xi_1, \xi_2, \cdots$ 独立同分布随机变量序列, 且 $\mathbf{E}|\xi_1| > 0$, 证明

$$\overline{\lim_{n\to\infty}} \left|\sum_{k=1}^n \xi_k\right| = +\infty \quad (\mathbf{P} - \text{a.c.}).$$

§2. 级数的收敛性

1. 独立随机变量的级数的敛散性准则　假设 $\xi_1, \xi_2, \cdots$ 是独立随机变量序列, $S_n = \xi_1 + \cdots + \xi_n$, 而 A 是使级数 $\sum \xi_n(\omega)$ 收敛于有限极限的基本事件 ω 的集合. 由柯尔莫戈洛夫 0–1 律知, 概率 $\mathbf{P}(A) = 0$ 或 1, 即级数 $\sum \xi_n$ 以概率 1 收敛或发散. 这一节的目的是, 给出确定独立随机变量的级数收敛还是发散的准则.

定理 1 (柯尔莫戈洛夫 – 辛钦)　a) 设 $\mathbf{E}\xi_n = 0, n \geqslant 1$. 那么, 如果

$$\sum \mathbf{E}\xi_n^2 < \infty, \tag{1}$$

则级数 $\sum \xi_n$ 以概率 1 收敛.

b) 并且, 如果随机变量 $\xi_n, n \geqslant 1$, 一致有界 (对于某个 $c < \infty, \mathbf{P}\{|\xi_n| \leqslant c\} = 1$), 则逆命题也成立: 若级数 $\sum \xi_n$ 以概率 1 收敛, 则条件 (1) 成立.

该定理的证明本质上用到下面的不等式.

柯尔莫戈洛夫不等式　a) 设 $\xi_1, \xi_2, \cdots, \xi_n$ 是独立随机变量, $\mathbf{E}\xi_i = 0, \mathbf{E}\xi_i^2 < \infty, 0 \leqslant i \leqslant n$. 那么, 对于任意 $\varepsilon > 0$, 有

$$\mathbf{P}\left\{\max_{1\leqslant k\leqslant n} |S_k| \geqslant \varepsilon\right\} \leqslant \frac{\mathbf{E}S_n^2}{\varepsilon^2}. \tag{2}$$

b) 并且, 如果 $\mathbf{P}\{|\xi_i| \leqslant c\} = 1, 0 \leqslant i \leqslant n$, 则

$$\mathbf{P}\left\{\max_{1\leqslant k\leqslant n} |S_k| \geqslant \varepsilon\right\} \geqslant 1 - \frac{(c+\varepsilon)^2}{\mathbf{E}S_n^2}. \tag{3}$$

证明 a) 记

$$A = \left\{ \max_{1 \leqslant k \leqslant n} |S_k| \geqslant \varepsilon \right\},$$

$$A_k = \{|S_i| < \varepsilon, i = 1, \cdots, k-1, |S_k| \geqslant \varepsilon\}, \quad 1 \leqslant k \leqslant n.$$

那么, $A = \sum A_k$, 而

$$\mathbf{E}S_n^2 \geqslant \mathbf{E}S_n^2 I_A = \sum \mathbf{E}S_n^2 I_{A_k}.$$

因为

$$\begin{aligned}\mathbf{E}S_n^2 I_{A_k} &= \mathbf{E}[S_k + (\xi_{k+1} + \cdots + \xi_n)]^2 I_{A_k} \\ &= \mathbf{E}S_k^2 I_{A_k} + 2\mathbf{E}S_k(\xi_{k+1} + \cdots + \xi_n) I_{A_k} + \mathbf{E}(\xi_{k+1} + \cdots + \xi_n)^2 I_{A_k} \geqslant \mathbf{E}S_k^2 I_{A_k},\end{aligned}$$

其中由于独立性假设和条件 $\mathbf{E}\xi_i = 0, 0 \leqslant i \leqslant n$, 可见

$$\mathbf{E}S_k(\xi_{k+1} + \cdots + \xi_n) I_{A_k} = \mathbf{E}S_k I_{A_k} \times \mathbf{E}(\xi_{k+1} + \cdots + \xi_n) = 0,$$

所以

$$\mathbf{E}S_n^2 \geqslant \sum \mathbf{E}S_k^2 I_{A_k} \geqslant \varepsilon^2 \sum \mathbf{P}(A_k) = \varepsilon^2 \mathbf{P}(A),$$

于是, 不等式 (2) 得证.

为证明不等式 (3), 注意到

$$\mathbf{E}S_n^2 I_A = \mathbf{E}S_n^2 - \mathbf{E}S_n^2 I_{\overline{A}} \geqslant \mathbf{E}S_n^2 - \varepsilon^2 \mathbf{P}(\overline{A}) = \mathbf{E}S_n^2 - \varepsilon^2 + \varepsilon^2 \mathbf{P}(A). \tag{4}$$

另一方面, 在集合 A_k 上, 有

$$|S_{k-1}| \leqslant \varepsilon, \quad |S_k| \leqslant |S_{k-1}| + |\xi_k| \leqslant \varepsilon + c,$$

因而

$$\begin{aligned}\mathbf{E}S_n^2 I_{A_n} &= \sum_k \mathbf{E}S_k^2 I_{A_k} + \sum_k \mathbf{E}[I_{A_k}(S_n - S_k)^2] \\ &\leqslant (\varepsilon + c)^2 \sum_k \mathbf{P}(A_k) + \sum_{k=1}^n \mathbf{P}(A_k) \sum_{j=k+1}^n \mathbf{E}\xi_j^2 \\ &\leqslant \mathbf{P}(A) \left[(\varepsilon + c)^2 + \sum_{j=1}^n \mathbf{E}\xi_j^2 \right] = \mathbf{P}(A)[(\varepsilon + c)^2 + \mathbf{E}S_n^2].\end{aligned} \tag{5}$$

由式 (4) 和 (5), 得

$$\mathbf{P}(A) \geqslant \frac{\mathbf{E}S_n^2 - \varepsilon^2}{(\varepsilon + c)^2 + \mathbf{E}S_n^2 - \varepsilon^2} = 1 - \frac{(\varepsilon + c)^2}{(\varepsilon + c)^2 + \mathbf{E}S_n^2 - \varepsilon^2} \geqslant 1 - \frac{(\varepsilon + c)^2}{\mathbf{E}S_n^2}.$$

于是, 不等式 (3) 得证. □

定理 1 的证明 a) 根据第二章 §10 定理 4, 序列 $(S_n)_{n\geqslant 1}$ 以概率 1 收敛, 当且仅当该序列以概率 1 是基本序列. 根据第二章 §10 定理 1, 序列 $(S_n)_{n\geqslant 1}$为基本序列 $(\mathbf{P}-\text{a.c.})$ 的充分必要条件是, 对于任意 $\varepsilon>0$, 有

$$\mathbf{P}\left\{\sup_{k\geqslant 1}|S_{n+k}-S_n|\geqslant\varepsilon\right\}\to 0,\quad n\to\infty. \tag{6}$$

由 (2) 式, 有

$$\mathbf{P}\left\{\sup_{k\geqslant 1}|S_{n+k}-S_n|\geqslant\varepsilon\right\}=\lim_{N\to\infty}\mathbf{P}\left\{\max_{1\leqslant k\leqslant N}|S_{n+k}-S_n|\geqslant\varepsilon\right\}$$
$$\leqslant\lim_{N\to\infty}\frac{\sum\limits_{k=n}^{n+N}\mathbf{E}\xi_k^2}{\varepsilon^2}=\frac{\sum\limits_{k=n}^{\infty}\mathbf{E}\xi_k^2}{\varepsilon^2}.$$

因此, 如果 $\sum\limits_{k=1}^{\infty}\mathbf{E}\xi_k^2<\infty$, 则条件 (6) 成立, 于是级数 $\sum\xi_k$ 以概率 1 收敛.

b) 设级数 $\sum\xi_k$ 以概率 1 收敛, 则由 (6) 式知, 对于充分大的 n, 有

$$\mathbf{P}\left\{\sup_{k\geqslant 1}|S_{n+k}-S_n|\geqslant\varepsilon\right\}<\frac{1}{2}. \tag{7}$$

由 (3) 式, 有

$$\mathbf{P}\left\{\sup_{k\geqslant 1}|S_{n+k}-S_n|\geqslant\varepsilon\right\}\geqslant 1-\frac{(c+\varepsilon)^2}{\sum\limits_{k=n}^{\infty}\mathbf{E}\xi_k^2}.$$

从而, 如果满足 $\sum\limits_{k=1}^{\infty}\mathbf{E}\xi_k^2=\infty$, 则

$$\mathbf{P}\left\{\sup_{k\geqslant 1}|S_{n+k}-S_n|\geqslant\varepsilon\right\}=1,$$

而这与 (7) 式矛盾. □

例 如果 $\xi_1,\xi_2,\cdots$ 是独立同分布伯努利随机变量序列, $\mathbf{P}\{\xi_n=+1\}=\mathbf{P}\{\xi_n=-1\}=1/2$, 则级数 $\sum\xi_n a_n(|a_n|\leqslant c)$ 以概率 1 收敛, 当且仅当 $\sum a_n^2<\infty$.

2. “两级数” 定理

定理 2 (柯尔莫戈洛夫 – 辛钦 “两级数” 定理) 独立随机变量的 $\xi_1,\xi_2,\cdots$ 的级数 $\sum\xi_n$ 以概率 1 收敛的充分条件是, 两级数 $\sum\mathbf{E}\xi_n$ 和 $\sum\mathbf{D}\xi_n$ 同时收敛. 并且, 如果对于某个 $c>0,\mathbf{P}\{|\xi_n|\leqslant c\}=1(n\geqslant 1)$, 则此条件也是必要的.

证明 如果 $\sum\mathbf{D}\xi_n<\infty$, 则根据定理 1, 级数 $\sum(\xi_n-\mathbf{E}\xi_n)$ 以概率 1 收敛. 因为根据假设级数 $\sum\mathbf{E}\xi_n$ 收敛, 所以级数 $\sum\xi_n$ 也以概率 1 收敛.

为证明必要性, 利用 “对称化” 方法. 与 $\xi_1,\xi_2,\cdots$ 同时, 考虑依赖于它的独立随机变量序列 $\widetilde{\xi}_1,\widetilde{\xi}_2,\cdots$, 其中 $\widetilde{\xi}_n$ 与 $\xi_n(n\geqslant 1)$ 同分布. (假如原基本事件空间是充分

"丰富" 的, 则由第二章 §9 定理 1 的系 1, 可见这样序列的存在性. 同时, 可以证明此假设并不失普遍性).

那么, 如果级数 $\sum \xi_n$ 以概率 1 收敛, 则级数 $\sum \widetilde{\xi}_n$ 以及级数 $\sum(\xi_n - \widetilde{\xi}_n)$ 也以概率 1 收敛. 而 $\mathbf{E}(\xi_n - \widetilde{\xi}_n) = 0$ 且 $\mathbf{P}\{|\xi_n - \widetilde{\xi}_n| \leqslant 2c\} = 1$. 所以由定理 1 的命题 b) 知, $\sum \mathbf{D}(\xi_n - \widetilde{\xi}_n) < \infty$. 此外,

$$\sum \mathbf{D}\xi_n = \frac{1}{2}\sum \mathbf{D}(\xi_n - \widetilde{\xi}_n) < \infty.$$

因此, 根据定理 1 的命题 a) , 级数 $\sum(\xi_n - \mathbf{E}\xi_n)$ 以概率 1 收敛, 说明级数 $\sum \mathbf{E}\xi_n$ 也收敛.

于是, 由于级数 $\sum \xi_n$ 以概率 1 收敛, 可见 (在假设 $\mathbf{P}\{|\xi_n| \leqslant c\} = 1 (n \geqslant 1)$ 下) 两级数 $\sum \mathbf{E}\xi_n$ 和 $\sum \mathbf{D}\xi_n$ 也同时收敛. □

3. "三级数" 定理

下面的定理, 在没有关于随机变量有界的假设条件下, 给出了级数 $\sum \xi_n$ 以概率 1 收敛的充分和必要条件.

对于某个常数 c, 设

$$\xi^c = \begin{cases} \xi, & 若\ |\xi| \leqslant c, \\ 0, & 若\ |\xi| > c. \end{cases}$$

定理 3 (柯尔莫戈洛夫 "三级数" 定理) *设 $\xi_1, \xi_2, \cdots$ 是独立随机变量序列. 级数 $\sum \xi_n$ 以概率 1 收敛的必要条件是, 对于任意 $c > 0$, 三级数*

$$\sum \mathbf{E}\xi_n^c, \quad \sum \mathbf{D}\xi_n^c, \quad \sum \mathbf{P}\{|\xi_n| \geqslant c\}$$

收敛, 并且三级数对于某个 $c > 0$ 收敛, 是级数 $\sum \xi_n^c$ 以概率 1 收敛的充分条件.

证明 (1) *充分性*. 根据 "两级数" 定理, 级数 $\sum \xi_n^c$ 以概率 1 收敛. 而如果对于某个 $c > 0, \sum \mathbf{P}\{|\xi_n| \geqslant c\} < \infty$, 则根据博雷尔 – 坎泰利引理, 以概率 1 级数 $\sum I\{|\xi_n| \geqslant c\} < \infty$, 说明对于一切 n (只可能有有限个可能除外), 有 $\xi_n = \xi_n^c$. 因此, 级数 $\sum \xi_n$ 也以概率 1 收敛.

(2) *必要性*. 如果级数 $\sum \xi_n$ 以概率 1 收敛, 则以概率 1 有 $\xi_n \to 0$, 故对于任意 $c > 0$, 以概率 1 最多有有限个事件 $\{|\xi_n| \geqslant c\}$ 出现. 因此, 以概率 1 有 $\sum I\{|\xi_n| \geqslant c\} < \infty$, 且根据第二博雷尔 – 坎泰利引理 $\sum \mathbf{P}\{|\xi_n| > c\} < \infty$. 此外, 由级数 $\sum \xi_n$ 收敛, 可见级数 $\sum \xi_n^c$ 收敛. 于是, 根据 "两级数" 定理, 两个级数 $\sum \mathbf{E}\xi_n^c$ 和 $\sum \mathbf{D}\xi_n^c$ 都收敛. □

系 *设 $\xi_1, \xi_2, \cdots$ 是独立随机变量, 且 $\mathbf{E}\xi_n = 0$. 那么, 如果*

$$\sum \mathbf{E}\frac{\xi_n^2}{1 + |\xi_n|} < \infty,$$

则级数 $\sum \xi_n$ 以概率 1 收敛.

为证明此结果, 我们注意到,

$$\sum \mathbf{E}\frac{\xi_n^2}{1+|\xi_n|} < \infty \Leftrightarrow \sum \mathbf{E}[\xi_n^2 I(|\xi_n| \leqslant 1) + |\xi_n| I(|\xi_n| > 1)] < \infty.$$

因此, 如果 $\xi_n^1 = \xi_n I(|\xi_n| \leqslant 1)$, 则

$$\sum \mathbf{E}(\xi_n^1)^2 < \infty.$$

由于 $\mathbf{E}\xi_n = 0$, 可见

$$\begin{aligned}\sum |\mathbf{E}\xi_n^1| = \sum \mathbf{E}|\xi_n I(|\xi_n| \leqslant 1)| &= \sum |\mathbf{E}\xi_n I(|\xi_n| > 1)| \\ \leqslant \sum \mathbf{E}|\xi_n| I(|\xi_n| > 1) &< \infty.\end{aligned}$$

说明级数 $\sum \mathbf{E}\xi_n^1$ 和 $\sum \mathbf{D}\xi_n^1$ 都收敛. 此外, 根据切比雪夫不等式, 有

$$\mathbf{P}\{|\xi_n| > 1\} = \mathbf{P}\{|\xi_n| I(|\xi_n| > 1)| > 1\} \leqslant \mathbf{E}|\xi_n| I(|\xi_n| > 1).$$

因此, $\sum \mathbf{P}\{|\xi_n| > 1\} < \infty$. 于是, 由 "三级数" 定理, 可见级数 $\sum \xi_n$ 以概率 1 收敛.

4. 练习题

1. 设 $\xi_1, \xi_2, \cdots$ 是独立随机变量序列, $S_n = \xi_1 + \cdots + \xi_n$. 利用 "三级数" 定理, 证明:

a) 如果级数 $\sum \xi_n^2$ 以概率 1 收敛, 则级数 $\sum \xi_n$ 以概率 1 收敛, 当且仅当级数 $\sum \mathbf{E}\xi_n I(|\xi_n| \leqslant 1)$ 收敛.

b) 如果级数 $\sum \xi_n$ 以概率 1 收敛, 则级数 $\sum \xi_n^2$ 也以概率 1 收敛, 当且仅当

$$\sum [\mathbf{E}|\xi_n| I(|\xi_n| \leqslant 1)]^2 < \infty.$$

2. 设 $\xi_1, \xi_2, \cdots$ 是独立随机变量序列, 证明级数 $\sum \xi_n^2$ 以概率 1 收敛, 当且仅当

$$\sum \mathbf{E}\frac{\xi_n^2}{1+\xi_n^2} < \infty.$$

3. 设 $\xi_1, \xi_2, \cdots$ 是独立随机变量序列, 证明如下三个条件等价:

a) 级数 $\sum \xi_n$ 以概率 1 收敛;

b) 级数 $\sum \xi_n$ 依概率收敛;

c) 级数 $\sum \xi_n$ 依分布收敛.

4. 举例说明, 在定理 1 和定理 2 中, 一般不能去掉一致有界性条件: 对于某个 $c > 0, \mathbf{P}\{|\xi_n| \leqslant c\} = 1 (n \geqslant 1)$.

5. 设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量, 且 $\mathbf{E}\xi_1 = 0, \mathbf{E}\xi_1^2 < \infty, S_n = \xi_1 + \cdots + \xi_n$. 证明如下柯尔莫戈洛夫不等式 (2) 的单侧类似 (马尔沙尔 [А. В. Маршалл]):

$$\mathbf{P}\left\{\max_{1\leqslant k\leqslant n} S_k \geqslant \varepsilon\right\} \leqslant \frac{\mathbf{E}S_n^2}{\varepsilon^2 + \mathbf{E}S_n^2}.$$

6. 设 $\xi_1, \xi_2, \cdots$ 是 (任意) 随机变量序列. 证明, 如果 $\sum_{n\geqslant 1} \mathbf{E}|\xi_n| < \infty$, 则级数 $\sum_{n\geqslant 1} \xi_n$ 以概率 1 绝对收敛.

7. 设 $\xi_1, \xi_2, \cdots$ 是独立对称分布随机变量. 证明

$$\mathbf{E}\left[\left(\sum_n \xi_n\right)^2 \wedge 1\right] \leqslant \sum_n \mathbf{E}(\xi_n^2 \wedge 1).$$

8. 设 $\xi_1, \xi_2, \cdots$ 是具有有限二阶矩的、独立随机变量. 证明, 当且仅当级数 $\sum \mathbf{E}\xi_n$ 和 $\sum \mathbf{D}\xi_n$ 收敛时, 级数 $\sum \xi_n$ 在 L^2 中收敛.

9. 设 $\xi_1, \xi_2, \cdots$ 是独立随机变量, 级数 $\sum \xi_n$ 几乎必然 (处处) 收敛. 证明, 该级数的值几乎必然与项的求和顺序无关, 当且仅当 $\sum |\mathbf{E}(\xi_n; |\xi_n| \leqslant 1)| < \infty$.

10. 设 $\xi_1, \xi_2, \cdots$ 是独立随机变量, 且 $\mathbf{E}\xi_n = 0\ (n \geqslant 1)$, 而

$$\sum_{n=1}^{\infty} \mathbf{E}[\xi_n^2 I(|\xi_n| \leqslant 1) + |\xi_n| I(|\xi_n| > 1)] < \infty.$$

证明级数 $\sum_{n=1}^{\infty} \xi_n$ $(\mathbf{P} - \text{a.c.})$ 收敛.

11. 设 $A_1, A_2, \cdots$ 是独立事件, 满足: $\mathbf{P}(A_n) > 0 (n \geqslant 1)$, $\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \infty$. 证明当 $n \to \infty$ 时, 有

$$\sum_{j=1}^{n} I(A_j) \Big/ \sum_{j=1}^{n} \mathbf{P}(A_j) \to 1 \quad (\mathbf{P} - \text{a.c.}).$$

12. 设 $\xi_1, \xi_2, \cdots$ 是独立随机变量, 其数学期望为 $\mathbf{E}\xi_n$ 和方差 σ_n^2 满足: $\lim_n \mathbf{E}\xi_n = c$, 和 $\sum_{n=1}^{\infty} \sigma_n^{-2} = \infty$. 证明当 $n \to \infty$ 时, 有

$$\sum_{j=1}^{n} \frac{\xi_j}{\sigma_j^2} \Big/ \sum_{j=1}^{n} \frac{1}{\sigma_j^2} \to c \quad (\mathbf{P} - \text{a.c.}).$$

§3. 强大数定律

1. 坎泰利强大数定律 设 $\xi_1, \xi_2, \cdots$ 是具有有限二阶矩的独立随机变量序列, $S_n = \xi_1 + \cdots + \xi_n$. 根据第三章 §3 练习题 2, 如果方差 $\mathbf{D}\xi_i$ 一致有界, 则大数定律成立:

$$\frac{S_n - \mathbf{E}S_n}{n} \xrightarrow{\mathbf{P}} 0, \quad n \to \infty. \tag{1}$$

强大数定律 如果将 (1) 式中的 "依概率收敛" 换成 "以概率 1 收敛", 则相应的论断称做强大数定律.

下面的定理, 是强大数定律最早的结果之一.

定理 1 (坎泰利) 设 $\xi_1, \xi_2, \cdots$ 是具有有限 4 阶矩的独立随机变量, 且对于某个常数 C, 有

$$\mathbf{E}|\xi_n - \mathbf{E}\xi_n|^4 \leqslant C, \quad n \geqslant 1.$$

那么, 当 $n \to \infty$时, 有

$$\frac{S_n - \mathbf{E}S_n}{n} \to 0, \quad (\mathbf{P} - \text{a.c.}). \tag{2}$$

证明 不失普遍性, 可以认为 $\mathbf{E}\xi_n = 0, n \geqslant 1$. 根据第二章 §10 定理 1 的系, 为使 $S_n/n \to 0$ ($\mathbf{P}$ – a.c.), 只需对于任何 $\varepsilon > 0$, 有

$$\sum \mathbf{P}\left\{\left|\frac{S_n}{n}\right| \geqslant \varepsilon\right\} < \infty.$$

同样地, 由切比雪夫不等式, 为满足上述不等式只需满足条件

$$\sum \mathbf{E}\left|\frac{S_n}{n}\right|^4 < \infty.$$

现在证明, 在定理的条件下该不等式确实成立.

易见,

$$S_n^4 = (\xi_1 + \cdots + \xi_n)^4 = \sum_{i=1}^n \xi_i^4 + \sum_{\substack{i,j\\ i<j}} \frac{4!}{2!2!}\xi_i^2\xi_j^2 + \sum_{\substack{i\neq j\\ i\neq k\\ j<k}} \frac{4!}{2!1!1!}\xi_i^2\xi_j\xi_k$$
$$+ \sum_{i<j<k<l} 4!\xi_i\xi_j\xi_k\xi_l + \sum_{i\neq j}\frac{4!}{3!1!}\xi_i^3\xi_j.$$

由于 $\mathbf{E}\xi_k = 0, k \leqslant n$, 故由此可见

$$\mathbf{E}S_n^4 = \sum_{i=1}^n \mathbf{E}\xi_i^4 + 6\sum_{i,j=1}^n \mathbf{E}\xi_i^2\mathbf{E}\xi_j^2 \leqslant nC + 6\sum_{\substack{i,j=1\\ i<j}}^n \sqrt{\mathbf{E}\xi_i^4\mathbf{E}\xi_j^4}$$
$$\leqslant nC + \frac{6n(n-1)}{2}C = (3n^2 - 2n)C < 3n^2C.$$

从而

$$\sum \mathbf{E}\left(\frac{S_n}{n}\right)^4 \leqslant 3C\sum\frac{1}{n^2} < \infty.$$ □

2. 强大数定律的柯尔莫戈洛夫准则 若引进更加精细的方法, 则可以本质上减弱定理 1 关于强大数定律成立的条件.

定理 2 (柯尔莫戈洛夫)[①] 设 $\xi_1, \xi_2, \cdots$ 是具有有限二阶矩的独立随机变量序列, 而 b_n 是满足如下条件的正数: $b_n \uparrow \infty$且

$$\sum \frac{\mathbf{D}\xi_n}{b_n^2} < \infty. \tag{3}$$

①该定理亦称做强大数定律的柯尔莫戈洛夫准则. —— 译者

那么，

$$\frac{S_n - \mathbf{E}S_n}{b_n} \to 0, \quad (\mathbf{P} - \text{a.c.}). \tag{4}$$

特别，如果

$$\sum \frac{\mathbf{D}\xi_n}{n^2} < \infty, \tag{5}$$

则

$$\frac{S_n - \mathbf{E}S_n}{n} \to 0, \quad (\mathbf{P} - \text{a.c.}). \tag{6}$$

定理 2 和下面定理 3 的证明，要用到两个辅助命题.

引理 1 (特普利茨 [O. Toeplitz]) 设 $(a_n)_{n\geqslant 1}$ 是非负数列，$b_n = \sum\limits_{i=1}^{n} a_i, b_1 = a_1 > 0$，且 $b_n \uparrow \infty, n \to \infty$；设 $(x_n)_{n\geqslant 1}$ 是收敛于某一实数 x 的数列. 那么，

$$\frac{1}{b_n}\sum_{j=1}^{n} a_j x_j \to x, \quad n \to \infty. \tag{7}$$

特别，如果 $a_n = 1$，则

$$\frac{x_1 + \cdots + x_n}{n} \to x, \quad n \to \infty. \tag{8}$$

证明 设 $\varepsilon > 0$ 和 $n_0 = n_0(\varepsilon)$ 满足：对于 $n \geqslant n_0, |x_n - x| \leqslant \varepsilon/2$. 选择 $n_1 > n_0$，使

$$\frac{1}{b_{n_1}}\sum_{j=1}^{n_0} a_j|x_j - x| < \frac{\varepsilon}{2}.$$

因此，对于 $n > n_1$，有

$$\begin{aligned}\left|\frac{1}{b_n}\sum_{j=1}^{n} a_j x_j - x\right| &\leqslant \frac{1}{b_n}\sum_{j=1}^{n} a_j|x_j - x| \\ &= \frac{1}{b_n}\sum_{j=1}^{n_0} a_j|x_j - x| + \frac{1}{b_n}\sum_{j=n_0+1}^{n} a_j|x_j - x| \\ &\leqslant \frac{1}{b_{n_1}}\sum_{j=1}^{n_0} a_j|x_j - x| + \frac{1}{b_n}\sum_{j=n_0+1}^{n} a_j|x_j - x| \\ &\leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2}\frac{b_n - b_{n_0}}{b_n} \leqslant \varepsilon.\end{aligned}$$

□

引理 2 (克罗内克 [L. Kronecker]) 设 $(b_n)_{n\geqslant 1}$ 是正实数的递增数列，并且 $b_n \uparrow \infty, n \to \infty$；而 $(x_n)_{n\geqslant 1}$ 是使级数 $\sum x_n$ 收敛的某一实数列. 那么，

$$\frac{1}{b_n}\sum_{j=1}^{n} b_j x_j \to 0, \quad n \to \infty.$$

特别，如果 $b_n = n, x_n = y_n/n$，而级数 $\sum y_n/n$ 收敛，则

$$\frac{y_1 + \cdots + y_n}{n} \to 0, \quad n \to \infty. \tag{9}$$

证明 设 $b_0 = 0, S_0 = 0, S_n = \sum_{j=1}^{n} x_j$. 那么 (“分部求和”), 有

$$\sum_{j=1}^{n} b_j x_j = \sum_{j=1}^{n} b_j(S_j - S_{j-1}) = b_n S_n - b_0 S_0 - \sum_{j=1}^{n} S_{j-1}(b_j - b_{j-1}),$$

设 $a_j = b_j - b_{j-1}$, 得

$$\frac{1}{b_n}\sum_{j=1}^{n} b_j x_j = S_n - \frac{1}{b_n}\sum_{j=1}^{n} S_{j-1} a_j \to 0,$$

因为若 $S_n \to x$, 则根据特普利茨引理, 有

$$\frac{1}{b_n}\sum_{j=1}^{n} S_{j-1} a_j \to x. \qquad \square$$

定理 2 的证明 因为

$$\frac{S_n - \mathbf{E}S_n}{b_n} = \frac{1}{b_n}\sum_{k=1}^{n} b_k \left(\frac{\xi_k - \mathbf{E}\xi_k}{b_k}\right),$$

则根据克罗内克引理, 为使 (4) 式成立, 只需级数 $\sum(\xi_k - \mathbf{E}\xi_k)/b_k$ 以概率 1 收敛. 实际上, 由于条件 (3) 和 §2 的定理 1 知, 该级数确实收敛.

例 1 设 $\xi_1, \xi_2, \cdots$ 是独立伯努利随机变量序列, 且 $\mathbf{P}\{\xi_n = 1\} = \mathbf{P}\{\xi_n = -1\} = 1/2$. 那么, 由

$$\sum \frac{1}{n \ln^2 n} < \infty, \quad \text{可见} \quad \frac{S_n}{\sqrt{n}\ln n} \to 0 \quad (\mathbf{P}-\text{a.c.}). \tag{10}$$

3. 独立同分布随机变量的强大数定律 对于随机变量 $\xi_1, \xi_2, \cdots$ 不但独立, 而且也同分布的情形, 为使强大数定律成立, 没有必要像定理 2 那样要求二阶矩存在, 只需要求一阶绝对矩存在.

定理 3 (柯尔莫戈洛夫) 设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量, 且 $\mathbf{E}|\xi_1| < \infty$. 那么,

$$\frac{S_n}{n} \to m \quad (\mathbf{P}-\text{a.c.}), \tag{11}$$

其中 $m = \mathbf{E}\xi_1, S_n = \xi_1 + \cdots + \xi_n$.

定理的证明用到下面的引理.

引理 3 设 ξ 是非负随机变量, 那么,

$$\sum_{n=1}^{\infty} \mathbf{P}\{\xi \geqslant n\} \leqslant \mathbf{E}\xi \leqslant 1 + \sum_{n=1}^{\infty} \mathbf{P}\{\xi \geqslant n\}. \tag{12}$$

引理 3 的证明. 下面的一系列不等式即可证明引理 3:

$$\sum_{n=1}^{\infty}\mathbf{P}\{\xi\geqslant n\}=\sum_{n=1}^{\infty}\sum_{k\geqslant n}\mathbf{P}\{k\leqslant\xi<k+1\}=\sum_{k=1}^{\infty}k\mathbf{P}\{k\leqslant\xi<k+1\}$$
$$=\sum_{k=0}^{\infty}\mathbf{E}[kI\{k\leqslant\xi<k+1\}]\leqslant\sum_{k=0}^{\infty}\mathbf{E}[\xi I\{k\leqslant\xi<k+1\}]=\mathbf{E}\xi$$
$$\leqslant\sum_{k=0}^{\infty}\mathbf{E}[(k+1)I\{k\leqslant\xi<k+1\}]=\sum_{k=0}^{\infty}(k+1)\mathbf{P}\{k\leqslant\xi<k+1\}$$
$$=\sum_{n=1}^{\infty}\mathbf{P}\{\xi\geqslant n\}+\sum_{k=0}^{\infty}\mathbf{P}\{k\leqslant\xi<k+1\}=\sum_{n=1}^{\infty}\mathbf{P}\{\xi\geqslant n\}+1.\qquad\square$$

定理 3 的证明 由引理 3 和博雷尔 - 坎泰利引理 (第二章 §10), 可见

$$\mathbf{E}|\xi_1|<\infty\Leftrightarrow\sum\mathbf{P}\{|\xi_1|\geqslant n\}<\infty\Leftrightarrow\sum\mathbf{P}\{|\xi_n|\geqslant n\}<\infty$$
$$\Leftrightarrow\sum\mathbf{P}\{\text{对于无限个 } n,|\xi_n|\geqslant n\}=0.$$

因此, 以概率 1 对于一切 n (仅有限个 n 有可能除外), 有 $|\xi_n|<n$.

记

$$\tilde{\xi}_n=\begin{cases}\xi_n, & |\xi_n|<n,\\ 0, & |\xi_n|\geqslant n,\end{cases}$$

且对于 $n\geqslant 1$, 认为 $\mathbf{E}\xi_n=0$. 那么,

$$\frac{\xi_1+\cdots+\xi_n}{n}\to 0,\quad n\to\infty(\mathbf{P}-\text{a.c.})\Leftrightarrow\frac{\tilde{\xi}_1+\cdots+\tilde{\xi}_n}{n}\to 0,\ n\to\infty(\mathbf{P}-\text{a.c.}).$$

注意, 一般 $\mathbf{E}\tilde{\xi}_n\neq 0$, 但是

$$\mathbf{E}\tilde{\xi}_n=\mathbf{E}\xi_nI\{|\xi_n|<n\}=\mathbf{E}\xi_1I\{|\xi_1|<n\}\to\mathbf{E}\xi_1=0.$$

因此, 由特普利茨引理, 有

$$\frac{1}{n}\sum_{k=1}^{n}\mathbf{E}\tilde{\xi}_k\to 0,\quad n\to\infty,$$

因而

$$\frac{\xi_1+\cdots+\xi_n}{n}\to 0,\quad n\to\infty(\mathbf{P}-\text{a.c.})$$
$$\Leftrightarrow\frac{(\tilde{\xi}_1-\mathbf{E}\tilde{\xi}_1)+\cdots+(\tilde{\xi}_n-\mathbf{E}\tilde{\xi}_n)}{n}\to 0,\ n\to\infty(\mathbf{P}-\text{a.c.}).\tag{13}$$

记 $\bar{\xi}_n=\tilde{\xi}_n-\mathbf{E}\tilde{\xi}_n$. 根据克罗内科引理, 为证明 (13) 式成立, 只需证明级数 $\sum\bar{\xi}_n/n$ 以概率 1 收敛. 而且由 §2 定理 1, 为此只需证明假设的条件 $\mathbf{E}|\xi_1|<\infty$ 可以保障级数 $\sum\mathbf{D}\bar{\xi}_n/n^2$ 收敛.

下面一系列不等式恰好证明这确实成立:

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{\mathbf{D}\bar{\xi}_n}{n^2} &\leqslant \sum_{n=1}^{\infty}\frac{\mathbf{E}\bar{\xi}_n^2}{n^2} = \sum_{n=1}^{\infty}\frac{1}{n^2}\mathbf{E}[\xi_n I(|\xi_n|<n)]^2\\ &= \sum_{n=1}^{\infty}\frac{1}{n^2}\mathbf{E}[\xi_1^2 I(|\xi_1|<n)] = \sum_{n=1}^{\infty}\frac{1}{n^2}\sum_{k=1}^{n}\mathbf{E}[\xi_1^2 I(k-1\leqslant|\xi_1|<k)]\\ &= \sum_{k=1}^{\infty}\mathbf{E}[\xi_1^2 I(k-1\leqslant|\xi_1|<k)]\sum_{n=k}^{\infty}\frac{1}{n^2} \leqslant 2\sum_{k=1}^{n}\frac{1}{k}\mathbf{E}[\xi_1^2 I(k-1\leqslant|\xi_1|<k)]\\ &\leqslant 2\sum_{k=1}^{\infty}\mathbf{E}[|\xi_1| I(k-1\leqslant|\xi_1|<k)] = 2\mathbf{E}|\xi_1|<\infty.\end{aligned}$$

□

注 1 在如下的意义上, 定理的逆命题也成立: 设 $\xi_1,\xi_2,\cdots$ 是独立同分布随机变量序列, 且以概率 1 满足

$$\frac{\xi_1+\cdots+\xi_n}{n}\to C,$$

其中 C 是某一 (有限) 常数. 那么, $\mathbf{E}|\xi_1|<\infty$ 且 $C=\mathbf{E}\xi_1$.

事实上, 如果 $S_n/n\to C(\mathbf{P}-\text{a.c.})$, 则

$$\frac{\xi_n}{n}=\frac{S_n}{n}-\left(\frac{n-1}{n}\right)\frac{S_{n-1}}{n-1}\to 0(\mathbf{P}-\text{a.c.}),$$

从而, $\mathbf{P}\{$对于无限个 $n, |\xi_n|>n\}=0$. 根据博雷尔 – 坎泰利引理 (第二章 §10), 且由引理 3 知 $\mathbf{E}|\xi_1|<\infty$, 有

$$\sum_{n=1}^{\infty}\mathbf{P}\{|\xi_1|>n\}<\infty.$$

那么, 由已经证明的定理, 可见 $C=\mathbf{E}\xi_1$.

于是, 对于独立同分布随机变量, 条件 $\mathbf{E}|\xi_1|<\infty$ 是 (以概率 1) 收敛于有限极限 S_n/n 的充分和必要条件.

注 2 如果数学期望 $m=\mathbf{E}\xi_1$ 存在, 但是未必有限, 则定理的断定 (9) 仍然成立.

事实上, 例如, 设 $\mathbf{E}\xi_1^-<\infty, \mathbf{E}\xi_1^+=\infty$. 对 $C>0$, 设

$$S_n^C=\sum_{i=1}^{n}\xi_i I(|\xi_i|\leqslant C).$$

那么, 以概率 1, 有

$$\varliminf_n\frac{S_n}{n}\geqslant\varliminf_n\frac{S_n^C}{n}=\mathbf{E}\xi_1 I(\xi_1\leqslant C).$$

由于当 $C\to\infty$ 时, 有

$$\mathbf{E}\xi_1 I(\xi_1\leqslant C)\to\mathbf{E}\xi_1=\infty,$$

可见 $S_n/n\to+\infty(\mathbf{P}-\text{a.c.})$.

注 3 定理 3 断定, $S_n/n \to m(\mathbf{P}-\text{a.c.})$. 需要指出, 这里除以概率 1 收敛之外, 平均收敛也成立:

$$\frac{S_n}{n} \xrightarrow{L^1} m, \quad \text{即} \quad \mathbf{E}\left|\frac{S_n}{n} - m\right| \to 0, \quad n \to \infty.$$

关于这一点, 由第五章 §3 的遍历性定理 3 的可以得到. 不过, 对于现在的独立同分布随机变量 $\xi_1, \xi_2, \cdots (S_n = \xi_1 + \cdots + \xi_n)$ 的情形, 可以直接证明 (练习题 7), 而无须借助于遍历性定理.

4. 应用强大数定律的例 现在考虑强大数定律的某些应用.

例 2 (用于数论) 设 $\Omega = [0,1), \mathscr{B}$ 是 Ω 的子集的 $\sigma-$ 代数, 而 $\mathbf{P}$ 是区间 $[0,1)$ 上的勒贝格测度. 考虑数 $\omega \in \Omega$ 的二进制分解 $\omega = 0.\ \omega_1\omega_2\cdots$ (含无限多个 0), 定义随机变量 $\xi_1(\omega), \xi_2(\omega), \cdots$, 其中 $\xi_n(\omega) = \omega_n$. 对于任意 $n \geqslant 1$ 和任意只取 0 或 1 为值的 $x_1, \cdots, x_n$, 有

$$\begin{aligned}
&\{\omega : \xi_1(\omega) = x_1, \cdots, \xi_n(\omega) = x_n\} \\
&= \left\{\omega : \frac{x_1}{2} + \frac{x_2}{2^2} + \cdots + \frac{x_n}{2^n} \leqslant \omega < \frac{x_1}{2} + \frac{x_2}{2^2} + \cdots + \frac{x_n}{2^n} + \frac{1}{2^n}\right\},
\end{aligned}$$

则该集合的 $\mathbf{P}-$ 测度等于 $1/2^n$. 说明 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量序列, 且

$$\mathbf{P}\{\xi_1 = 0\} = \mathbf{P}\{\xi_1 = 1\} = \frac{1}{2}.$$

由此和强大数定律得如下博雷尔的结果: 区间 $[0,1)$ 的几乎一切数按如下意义都是正规的: 其二进制分解中 "0 的比率" 和 "1 的比率" 各以概率 1 趋向 1/2, 即.

$$\frac{1}{n}\sum_{k=1}^{n} I(\xi_k = 1) \to \frac{1}{2} \quad (\mathbf{P}-\text{a.c.}).$$

例 3 (应用于 "蒙特卡罗法") 设 $f(x)$ 是定义在区间 [0,1] 上、取值于 [0,1] 的连续函数. 下面的讨论基于积分 $\int_0^1 f(x)\mathrm{d}x$ 的数值计算的统计方法 ("蒙特卡罗法").

设 $\xi_1, \eta_1, \xi_2, \eta_2, \cdots$ 是独立同在 [0,1] 上均匀分布的随机变量序列; 而

$$\rho_i = \begin{cases} 1, & \text{若 } f(\xi_i) > \eta_i, \\ 0, & \text{若 } f(\xi_i) \leqslant \eta_i. \end{cases}$$

显然,

$$\mathbf{E}\rho_1 = \mathbf{P}\{f(\xi_1) > \eta_1\} = \int_0^1 f(x)\mathrm{d}x.$$

由强大数定律 (定理 3), 有

$$\frac{1}{n}\sum_{i=1}^{n} \rho_i \to \int_0^1 f(x)\mathrm{d}x \quad (\mathbf{P}-\text{a.c.}).$$

这样, 积分 $\int_0^1 f(x)\mathrm{d}x$ 的数值计算, 可以利用随机数偶 $(\xi_i,\eta_i), i\geqslant 1$, 模拟, 再计算 ρ_i 和 $n^{-1}\sum\limits_{i=1}^{n}\rho_i$ 的值来实现.

例 4 (更新过程的强大数定律)　设 $N=(N_t)_{t\geqslant 0}$ 是第二章 §9 第 4 小节引进的更新过程:

$$N_t=\sum_{n=1}^{\infty} I(T_n\leqslant t),\quad T_n=\sigma_1+\cdots+\sigma_n,$$

其中 $\sigma_1,\sigma_2,\cdots,\sigma_n$ 是独立同分布正值随机变量序列. 此外, 假设 $\mu=\mathbf{E}\sigma_1<\infty$.

在这样的假设条件下, 过程 N 服从强大数定律:

$$\frac{N_t}{t}\to\frac{1}{\mu},\quad (\mathbf{P}-\text{a.c.}),\quad t\to\infty. \tag{14}$$

为证明这一结果, 首先由于 $T_{N_t}\leqslant t<T_{N_t+1}(t\geqslant 0)$, 可见在假设 $N_t>0$ 的条件下, 不等式

$$\frac{T_{N_t}}{N_t}\leqslant\frac{t}{N_t}<\frac{T_{N_t+1}}{N_t+1}\left(1+\frac{1}{N_t}\right) \tag{15}$$

成立.

显然, 当 $t\to\infty$ 时 $N_t=N_t(\omega)\to\infty$ $(\mathbf{P}-\text{a.c.})$. 同时, 根据定理 3, 当 $t\to\infty$ 时, 有

$$\frac{T_n(\omega)}{n}=\frac{\sigma_1(\omega)+\cdots+\sigma_n(\omega)}{n}\to\mu,\quad (\mathbf{P}-\text{a.c.}).$$

因此, 同样, 有

$$\frac{T_{N_t(\omega)}(\omega)}{N_t(\omega)}\to\mu,\quad (\mathbf{P}-\text{a.c.}),\quad n\to\infty,$$

从而, 由 (15) 式, 可见存在极限:

$$\lim_{t\to\infty}\frac{t}{N_t}=\mu\ \ (\mathbf{P}-\text{a.c.}).$$

于是, 强大数定律 (14) 得证.

5. 练习题

1. 证明 $\mathbf{E}\xi^2<\infty$, 当且仅当

$$\sum_{n=1}^{\infty} n\mathbf{P}\{|\xi|>n\}<\infty.$$

2. 假设 $\xi_1,\xi_2,\cdots$ 是独立同分布随机变量, 证明:

(a) 若对于某个 $0<\alpha<1$, 有 $\mathbf{E}|\xi_1|^\alpha<\infty$, 则 $S_n/n^{1/\alpha}\to 0(\mathbf{P}-\text{a.c.})$.

(b) 若对于某个 $1\leqslant\beta<2$, 有 $\mathbf{E}|\xi_1|^\beta<\infty$, 则

$$\frac{S_n-n\mathbf{E}\xi_1}{n^{1/\beta}}\to 0\quad (\mathbf{P}-\text{a.c.}).$$

3. 设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量序列, 且 $\mathbf{E}|\xi_1| = \infty$. 证明, 对于任意常数序列 $\{a_n\}$, 有

$$\overline{\lim_n} \left| \frac{S_n}{n} - a_n \right| = \infty \quad (\mathbf{P} - \text{a.c.}).$$

4. 问区间 $[0,1)$ 上的一切有理数 (在第 4 小节例 2) 是否都是正规的?

5. 举一独立随机变量序列 $\xi_1, \xi_2, \cdots$ 的例, 说明极限 $\lim\limits_n S_n/n$ 在依概率的意义下存在, 但是在以概率 1 的意义下不存在,

6. (埃特麦迪 [N. Etemady].) 证明, 如果将随机变量 $\xi_1, \xi_2, \cdots$ "独立" 换成 "两两独立", 定理 3 的结论仍然成立.

7. 证明. 在定理 3 的条件下, 平均收敛成立:

$$\mathbf{E} \left| \frac{S_n}{n} - m \right| \to 0, \quad n \to \infty.$$

8. 设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量序列, 且 $\mathbf{E}|\xi_1|^2 < \infty$. 证明,

$$n\mathbf{P}\{|\xi_1| \geqslant \varepsilon\sqrt{n}\} \to 0 \quad \text{和} \quad \frac{1}{\sqrt{n}} \max_{k \leqslant n} |\xi_k| \xrightarrow{\mathbf{P}} 0.$$

9. 考虑区间 $[0,1)$ 上数 $\omega = 0.\omega_1\omega_2\cdots$ 的十进制分解.

(a) 将第 4 小节对于二进制分解的强大数定律, 移植到上面二进制分解的情形.

(b) 证明, 有理数 (按博雷尔) 不是正规的, 即对于有理数的十进制分解 $(\xi_k(\omega) = \omega_k, k \geqslant 1)$, 对于任意 $i = 0, 1, \cdots, 9$,

$$\frac{1}{n} \sum_{k=1}^{n} I(\xi_k(\omega) = i) \nrightarrow \frac{1}{10} \quad (\mathbf{P} - \text{a.c.}).$$

(c) 证明, 数 $\omega = 0.12345678910111213\cdots$, 其中在小数位上依次写的是一切非负整数, 是正规的 (见例 2).

10. (a) 设 $\xi_1, \xi_2, \cdots$ 是独立随机变量序列, 且 $\mathbf{P}\{\xi_n = \pm n^\alpha\} = 1/2$. 证明, 这一随机变量列服从大数定律, 当且仅当 $\alpha < 1/2$.

(b) 设 $f = f(x)$ 是 $(0, \infty)$ 上的有界连续函数. 证明, 对于任何 $a > 0$ 与一切 $x > 0$, 有

$$\lim_{n \to \infty} \sum_{k=1}^{\infty} f\left(x + \frac{k}{n}\right) \mathrm{e}^{-an} \frac{(an)^k}{k!} = f(x+a).$$

11. 证明, 可以赋予柯尔莫戈洛夫强大数定律 (定理 3) 如下形式: 设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量, 则

$$\mathbf{E}|\xi_1| < \infty \Leftrightarrow n^{-1}S_n \to \mathbf{E}\xi_1 \quad (\mathbf{P} - \text{a.c.}),$$
$$\mathbf{E}|\xi_1| = \infty \Leftrightarrow \overline{\lim} n^{-1} S_n = +\infty \quad (\mathbf{P} - \text{a.c.}).$$

证明, 如果将 "独立" 换成 "两两独立", 则第一个结论仍然成立.

12. 设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量序列. 证明,

$$\mathbf{E}\sup_n\left|\frac{\xi_n}{n}\right| < \infty \Leftrightarrow \mathbf{E}|\xi_1|\ln^+|\xi_1| < \infty.$$

13. 设 $S_n = \xi_1 + \cdots + \xi_n, n \geqslant 1$, 其中假设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量序列, 且 $\mathbf{E}\xi_1 = 0, \mathbf{E}|\xi_1| > 0$. 证明

$$\overline{\lim_n}\frac{S_n}{\sqrt{n}} = \infty, \quad \underline{\lim}\frac{S_n}{\sqrt{n}} = -\infty \quad (\mathbf{P} - \text{a.c.}).$$

14. 设 $S_n = \xi_1 + \cdots + \xi_n, n \geqslant 1$, 其中设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量序列. 证明. 对于任意 $\alpha \in (0, 1/2]$, 下面的性质之一成立:

(a) $n^{-\alpha}S_n \to \infty(\mathbf{P} - \text{a.c.})$;

(b) $n^{-\alpha}S_n \to -\infty(\mathbf{P} - \text{a.c.})$;

(c) $\overline{\lim}n^{-\alpha}S_n = \infty, \underline{\lim}n^{-\alpha}S_n = -\infty, (\mathbf{P} - \text{a.c.})$.

15. 设 $S_n = \xi_1 + \cdots + \xi_n, n \geqslant 1, S_0 = 0, \xi_1, \xi_2, \cdots$ 是独立同分布随机变量序列. 证明:

(a) 对于任意 $\varepsilon > 0$,

$$\sum_{n=1}^{\infty}\mathbf{P}\{|S_n| \geqslant n\varepsilon\} < \infty \Leftrightarrow \mathbf{E}\xi_1 = 0, \quad \mathbf{E}\xi_1^2 < \infty;$$

(b) 如果 $\mathbf{E}\xi_1 < 0$, 则对于 $p > 1$,

$$\mathbf{E}\left(\sup_{n\geqslant 0} S_n\right)^{p-1} < \infty \Leftrightarrow \mathbf{E}(\xi_1^+)^p < \infty;$$

(c) 如果 $\mathbf{E}\xi_1 = 0, 1 < p \leqslant 2$, 则对于某个常数 C_p,

$$\sum_{n=1}^{\infty}\mathbf{P}\left\{\max_{k\leqslant n} S_k \geqslant n\right\} \leqslant C_p\mathbf{E}|\xi_1|^p, \quad \sum_{n=1}^{\infty}\mathbf{P}\left\{\max_{k\leqslant n}|S_k| \geqslant n\right\} \leqslant 2C_p\mathbf{E}|\xi_1|^p;$$

(d) 如果 $\mathbf{E}\xi_1 = 0, \mathbf{E}\xi_1^2 < \infty$ 且 $M(\varepsilon) = \sup\limits_{n\geqslant 0}(S_n - n\varepsilon), \varepsilon > 0$, 则

$$\lim_{\varepsilon\to 0}\varepsilon M(\varepsilon) = \frac{\sigma^2}{2}.$$

§4. 重对数定律

1. 辅助函数: 上函数和下函数 设 $\xi_1, \xi_2, \cdots$ 是独立同分布伯努利随机变量序列, 其中 $\mathbf{P}\{\xi_n = 1\} = \mathbf{P}\{\xi_n = -1\} = 1/2, S_n = \xi_1 + \cdots + \xi_n$. 由 §1 定理 2 的证明, 可见

$$\overline{\lim}\frac{S_n}{\sqrt{n}} = +\infty, \quad \underline{\lim}\frac{S_n}{\sqrt{n}} = -\infty. \tag{1}$$

另一方面, 根据 §3 的 (10) 式, 有

$$\frac{S_n}{\sqrt{n}\ln n} \to 0 \quad (\mathbf{P}-\text{a.c.}). \tag{2}$$

现在比较这两个结果.

由 (1) 式可见, 轨道 $(S_n)_{n\geqslant 1}$ 以概率 1 无限多次与 "曲线" $\pm\varepsilon\sqrt{n}$ 相交, 其中 ε 是任意正数. 但是, 与此同时由 (2) 式可见, 只有有限多次越出有界曲线 $\pm\varepsilon\sqrt{n}\ln n$ 所围成的区域内部. 这两个关于对称随机游动 $(S_n)_{n\geqslant 1}$ 振荡 "幅度" 特点的结果, 提供了相当有用的信息. 下面引进的重对数定律, 本质上更精确地说明了关于随机游动 $(S_n)_{n\geqslant 1}$ 振荡 "幅度" 的概念.

首先引进上函数和下函数的概念.

定义 函数 $\varphi^* = \varphi^*(n), n \geqslant 1$, (对于 $(S_n)_{n\geqslant 1}$) 称做**上函数**, 如果自某个 $n = n_0(\omega)$ 开始对于一切 n, 以概率 1 有 $S_n \leqslant \varphi^*(n)$.

函数 $\varphi_* = \varphi_*(n), n \geqslant 1$ (对于 $(S_n)_{n\geqslant 1}$) 称做**下函数**, 如果对无限多个 n, 以概率 1 有 $S_n > \varphi_*(n)$.

根据这一定义, 由 (1) 式和 (2) 式可见, 每一个函数 $\varphi^* = \varepsilon\sqrt{n}\ln n(\varepsilon > 0)$, (对于 $(S_n)_{n\geqslant 1}$) 是**上函数**, 而函数 $\varphi_* = \varepsilon\sqrt{n}(\varepsilon > 0)$ 是**下函数**.

设 $\varphi = \varphi(n)$ 是某个函数, 而 $\varphi_\varepsilon^* = (1+\varepsilon)\varphi, \varphi_{*\varepsilon} = (1-\varepsilon)\varphi$, 其中 $\varepsilon > 0$. 那么, 易见

$$\begin{aligned}
\left\{\overline{\lim_n}\frac{S_n}{\varphi(n)} \leqslant 1\right\} &= \left\{\lim_n\left[\sup_{m\geqslant n}\frac{S_m}{\varphi(m)}\right] \leqslant 1\right\} \\
&\Leftrightarrow \left\{\sup_{m\geqslant n_1(\varepsilon)}\frac{S_m}{\varphi(m)} \leqslant 1+\varepsilon,\ \text{对于一切 } \varepsilon > 0 \text{ 和某个 } n_1(\varepsilon)\right\} \\
&\Leftrightarrow \{S_m \leqslant (1+\varepsilon)\varphi(m) \text{ 对一切 } \varepsilon > 0 \text{ 和自某个 } n_1(\varepsilon) \text{ 始的一切 } m\}.
\end{aligned} \tag{3}$$

同样, 有

$$\begin{aligned}
\left\{\overline{\lim_n}\frac{S_n}{\varphi(n)} \geqslant 1\right\} &= \left\{\lim_n\left[\sup_{m\geqslant n}\frac{S_m}{\varphi(m)}\right] \geqslant 1\right\} \\
&\Leftrightarrow \left\{\sup_{m\geqslant n_2(\varepsilon)}\frac{S_m}{\varphi(m)} \geqslant 1-\varepsilon,\ \text{对于一切 } \varepsilon > 0 \text{ 和某个 } n_2(\varepsilon)\right\} \\
&\Leftrightarrow \{S_m \geqslant (1-\varepsilon)\varphi(m) \text{ 对一切 } \varepsilon > 0 \text{ 和自某个 } n_3(\varepsilon) \geqslant n_2(\varepsilon) \text{ 始的无限个 } m\}.
\end{aligned} \tag{4}$$

由 (3) 式和 (4) 式可见, 为验证每一个函数 $\varphi_\varepsilon^* = (1+\varepsilon)\varphi, \varepsilon > 0$, 是上函数, 需要证明:

$$\mathbf{P}\left\{\overline{\lim_n}\frac{S_n}{\varphi(n)} \leqslant 1\right\} = 1. \tag{5}$$

而为证明每一个函数 $\varphi_{*\varepsilon} = (1-\varepsilon)\varphi, \varepsilon > 0$ 是下函数, 需要验证:

$$\mathbf{P}\left\{\overline{\lim_n}\frac{S_n}{\varphi(n)} \geqslant 1\right\} = 1. \tag{6}$$

2. 重对数定律

定理 1 (重对数定律) 设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量序列, 且 $\mathbf{E}\xi_i = 0$, $\mathbf{E}\xi_i^2 = \sigma^2 > 0$. 那么,

$$\mathbf{P}\left\{\overline{\lim_n} \frac{S_n}{\psi(n)} = 1\right\} = 1, \tag{7}$$

其中

$$\psi(n) = \sqrt{2\sigma^2 n \ln\ln n}. \tag{8}$$

对于一致有界随机变量, 1924 年由辛钦证明了重对数定律. 1929 年柯尔莫戈洛夫将这一结果, 推广到广泛的独立随机变量类. 在定理 1 所表述的条件下, 1941 年由哈特曼 (P. Hartman) 和温特纳 (A. Wintner) 证明了重对数定律.

由于该定理的证明相当复杂, 故我们仅限于考虑其特殊情形, 即随机变量 ξ_n 服从正态分布的情形: $\xi_n \sim N(0,1), n \geqslant 1$.

首先, 证明两个引理.

引理 1 设 $\xi_1, \cdots, \xi_n$ 是独立随机变量, 且有对称分布: 对于每一个事件 $B \in \mathscr{B}(\mathbb{R}), k \leqslant n$, 有 $\mathbf{P}\{\xi_k \in B\} = \mathbf{P}\{-\xi_k \in \mathbf{B}\}$. 那么, 对于任意实数 a, 有

$$\mathbf{P}\left\{\max_{1\leqslant k\leqslant n} S_k > a\right\} \leqslant 2\mathbf{P}\{S_n > a\}. \tag{9}$$

证明 引进记号: $A_k = \{S_i \leqslant a, i \leqslant k-1; S_k > a\}, A = \left\{\max_{1\leqslant k\leqslant n} S_k > a\right\}, B = \{S_n > a\}$. 由于在 $A_k \bigcap B \supseteq A_k \bigcap \{S_n \geqslant S_k\}$ 上, 则

$$\begin{aligned}\mathbf{P}(B\bigcap A_k) &\geqslant \mathbf{P}(A_k\bigcap\{S_n \geqslant S_k\}) = \mathbf{P}(A_k)\mathbf{P}\{S_n \geqslant S_k\} \\ &= \mathbf{P}(A_k)\mathbf{P}\{\xi_{k+1} + \cdots + \xi_n \geqslant 0\}.\end{aligned}$$

由于随机变量 $\xi_1, \cdots, \xi_n$ 的概率分布对称, 有

$$\mathbf{P}\{\xi_{k+1} + \cdots + \xi_n > 0\} = \mathbf{P}\{\xi_{k+1} + \cdots + \xi_n < 0\}.$$

因此 $\mathbf{P}\{\xi_{k+1} + \cdots + \xi_n \geqslant 0\} \geqslant 1/2$, 从而

$$\mathbf{P}(B) \geqslant \sum_{k=1}^n \mathbf{P}(A_k\bigcap B) \geqslant \frac{1}{2}\sum_{k=1}^n \mathbf{P}(A_k) = \frac{1}{2}\mathbf{P}(A),$$

于是 (9) 式得证. (对照第八章 §2 第 3 小节的证明.) □

引理 2 设 $S_n \sim N(0, \sigma^2(n)), \sigma^2(n) \uparrow \infty$, 而数 $a(n), n \geqslant 1$, 满足 $a(n)/\sigma(n) \to \infty, n \to \infty$. 那么

$$\mathbf{P}\{S_n > a(n)\} \sim \frac{\sigma(n)}{\sqrt{2\pi}a(n)} \mathrm{e}^{-\frac{a^2(n)}{2\sigma^2(n)}}. \tag{10}$$

证明 由于当 $x \to \infty$ 时

$$\frac{1}{\sqrt{2\pi}} \int_x^{\infty} \mathrm{e}^{-\frac{y^2}{2}} \mathrm{d}y \sim \frac{1}{\sqrt{2\pi} x} \mathrm{e}^{-\frac{x^2}{2}},$$

而 $S_n/\sigma^2(n) \sim N(0,1)$, 立即可得 (10) 式.

证明定理 1 现在证明 $\xi \sim N(0,1)$ 的情形. 假设 $\xi_i \sim N(0,1)$.

(a) 首先证明 (5) 式. 设 $\varepsilon > 0, \lambda = 1+\varepsilon, n_k = \lambda^k$, 其中 $k \geqslant k_0$, 且选择 k_0, 使 $\ln\ln k_0$ 有定义. 记

$$A_k = \{S_n > \lambda\psi(n),\ 对于某个\ n \in (n_k, n_{k+1}]\}, \tag{11}$$

亦设

$$A = \{A_k,\ 对无限多个\ k\} = \{S_n > \lambda\psi(n),\ 对于无限多个\ n\}.$$

由于 (3) 式, 为证明 (5) 式, 只需证明 $\mathbf{P}(A) = 0$.

为此证明 $\sum \mathbf{P}(A_k) < \infty$. 根据博雷尔 – 坎泰利引理 (第二章 §10), 有 $\mathbf{P}(A) = 0$.

由 (11),(9) 和 (10) 式, 有

$$\begin{aligned}
&\mathbf{P}(A_k) \leqslant \mathbf{P}\{S_n > \lambda\psi(n_k),\ 对于某个\ n \in (n_k, n_{k+1}]\} \\
&\leqslant \mathbf{P}\{S_n > \lambda\psi(n_k),\ 对于某个\ n \leqslant n_{k+1}\} \leqslant 2\mathbf{P}\{S_{n_{k+1}} > \lambda\psi(n_k)\} \\
&\sim \frac{2}{\sqrt{2\pi}\dfrac{\lambda\psi(n_k)}{\sqrt{n_k}}} \mathrm{e}^{-\frac{1}{2}\left(\frac{\lambda\psi(n_k)}{\sqrt{n_k}}\right)^2} \leqslant C_1 \mathrm{e}^{-\lambda \ln\ln \lambda^k} \leqslant C_2 \mathrm{e}^{-\lambda \ln k} = C_2 k^{-\lambda},
\end{aligned}$$

其中 C_1 和 C_2 是常数. 由于 $\sum\limits_{k=1}^{\infty} k^{-\lambda} < \infty$, 故 $\sum \mathbf{P}(A_k) < \infty$.

这样, (5) 式得证.

(b) 其次, 证明 (6) 式. 根据 (4) 式, 需要证明对于 $\lambda = 1-\varepsilon, \varepsilon > 0$, 以概率 1 对无限多个 n, 有 $S_n \geqslant \lambda\psi(n)$. 现在将已证明的 (5) 式用于序列 $(-S_n)_{n\geqslant 1}$, 则有可能除有限个 n 的值之外的一切 n, ($\mathbf{P}$ – a.c.), 有 $-S_n \leqslant 2\psi(n)$. 从而, 如果 $n_k = N^k, N > 1$, 则对于充分大的 k, 有

$$S_{n_{k-1}} \geqslant -2\psi(n_{k-1}),$$

或

$$S_{n_k} \geqslant Y_k - 2\psi(n_{k-1}), \tag{12}$$

其中 $Y_k = S_{n_k} - S_{n_{k-1}}$.

因此, 如果证明了, 对无限多个 k,

$$Y_k > \lambda\psi(n_k) + 2\psi(n_{k-1}), \tag{13}$$

则这连同 (12) 式可以证明, 对于无限多个 k, 以概率 1 有 $S_{n_k} > \lambda\psi(n_k)$. 取某个 $\lambda' \in (\lambda, 1)$. 那么, 存在 $N > 1$, 使对一切 k,

$$\begin{aligned} &\lambda'[2(N^k - N^{k-1}) \ln\ln N^k]^{1/2} \\ &> \lambda(2N^k \ln\ln N^k)^{1/2} + 2(2N^{k-1} \ln\ln N^{k-1})^{1/2} \\ &\equiv \lambda\psi(N^k) + 2\psi(N^{k-1}). \end{aligned}$$

现在只需证明, 对于无限多个 k,

$$Y_k > \lambda'[2(N^k - N^{k-1}) \ln\ln N^k]^{1/2}. \tag{14}$$

显然, $Y_k \sim N(0, N^k - N^{k-1})$. 因此, 由于引理 2, 可见

$$\begin{aligned} &\mathbf{P}\{Y_k > \lambda'[2(N^k - N^{k-1}) \ln\ln N^k]^{1/2}\} \\ &\sim \frac{1}{\sqrt{2\pi}\lambda'(2\ln\ln N^k)^{1/2}} \mathrm{e}^{(-\lambda')^2 \ln\ln N^k} \\ &\geqslant \frac{C_1}{(\ln k)^{1/2}} k^{-(\lambda')^2} \geqslant \frac{C_2}{k \ln k}. \end{aligned}$$

因为 $\sum (k \ln k)^{-1} = \infty$, 则根据博雷尔 – 坎泰利引理的第 2 部分, 以概率 1 对于无限多个 k, (14) 式成立. 于是关系式 (6) 得证.

(c) 由 (5) 式和 (6) 式, 立即得所要证明的 (7) 式. □

注 1　将 (7) 式用于随机变量 $(-S_n)_{n\geqslant 1}$, 得 ($\mathbf{P}$ – a.c.)

$$\underline{\lim} \frac{S_n}{\psi(n)} = -1. \tag{15}$$

由 (7) 和 (15) 式可见, 重对数定律亦可表示为如下形式:

$$\mathbf{P}\left\{\overline{\lim} \frac{S_n}{\psi(n)} = 1\right\} = 1. \tag{16}$$

注 2　重对数定律说明, 对于任意 $\varepsilon > 0$, 每一个函数 $\psi_\varepsilon^* = (1+\varepsilon)\psi$ 都是上函数, 而函数 $\psi_{*\varepsilon} = (1-\varepsilon)\psi$ 是下函数.

重对数定律的结论 7 与下面的关系式等价: 对于任意 $\varepsilon > 0$,

$$\begin{aligned} &\mathbf{P}\{|S_n| \geqslant (1-\varepsilon)\psi(n),\ \ \text{对于无限多个 } n\} = 1, \\ &\mathbf{P}\{|S_n| \geqslant (1+\varepsilon)\psi(n),\ \ \text{对于无限多个 } n\} = 0. \end{aligned}$$

3. 练习题

1. 设 $\xi_1, \xi_2, \cdots$ 是独立随机变量序列, 且 $\xi_n \sim N(0,1)$. 证明

$$\mathbf{P}\left\{\overline{\lim_n} \frac{\xi_n}{\sqrt{2\ln n}} = 1\right\} = 1.$$

2. 设 $\xi_1, \xi_2, \cdots$ 是独立随机变量, 服从参数为 $\lambda > 0$ 的泊松分布. 证明 (与 λ 无关)

$$\mathbf{P}\left\{\overline{\lim_n}\frac{\xi_n \ln\ln n}{\ln n} = 1\right\} = 1.$$

3. 设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量序列, 其共同的特征函数为

$$\mathbf{E}\mathrm{e}^{-it\xi_1} = \mathrm{e}^{-|t|^\alpha}, \quad 0 < \alpha < 2.$$

证明

$$\mathbf{P}\left\{\overline{\lim_n}\left|\frac{S_n}{n^{1/\alpha}}\right|^{\frac{1}{\ln\ln n}} = \mathrm{e}^{1/\alpha}\right\} = 1.$$

4. 证明不等式 (9) 的如下推广的正确性. 设 $\xi_1, \cdots, \xi_n$ 是独立随机变量, $S_0 = 0, S_k = \xi_1 + \cdots + \xi_k$, 则对于任意实数 a, 列维 (P. P. Lévy) 不等式成立:

$$\mathbf{P}\left\{\max_{1\leqslant k\leqslant n}[S_k + \mu(S_n - S_k)] > a\right\} \leqslant 2\mathbf{P}\{S_n > a\},$$

其中 $\mu(\xi)$ 是随机变量 ξ 的中位数, 即满足下列不等式的常数:

$$\mathbf{P}\{\xi \geqslant \mu(\xi)\} \geqslant \frac{1}{2}, \quad \mathbf{P}\{\xi \leqslant \mu(\xi)\} \geqslant \frac{1}{2}.$$

5. 设 $\xi_1, \cdots, \xi_n$ 是独立随机变量, $S_0 = 0, S_k = \xi_1 + \cdots + \xi_k$, 证明

(a) (第 4 题的补充)

$$\mathbf{P}\left\{\max_{1\leqslant k\leqslant n}|S_k + \mu(S_n - S_k)| > a\right\} \leqslant 2\mathbf{P}\{|S_n| \geqslant a\},$$

其中 $\mu(\xi)$ 是随机变量 ξ 的中位数.

(b) 如果 $\xi_1, \cdots, \xi_n$ 是同分布对称随机变量, 则

$$1 - \mathrm{e}^{-n\mathbf{P}\{|\xi_1| > x\}} \leqslant \mathbf{P}\left\{\max_{1\leqslant k\leqslant n}|\xi_k| > x\right\} \leqslant 2\mathbf{P}\{|S_n| > x\}.$$

6. 设 $\xi_1, \cdots, \xi_n$ 是独立随机变量, 且 $\mathbf{E}\xi_i = 0 (1 \leqslant i \leqslant n), S_k = \xi_1 + \cdots + \xi_k$, 证明对于 $a > 0$,

$$\mathbf{P}\left\{\max_{1\leqslant k\leqslant n} S_k > a\right\} \leqslant 2\mathbf{P}\{S_n \geqslant a - \mathbf{E}|S_n|\}.$$

7. 设 $\xi_1, \cdots, \xi_n$是独立同分布随机变量, 且 $\mathbf{E}\xi_i = 0, \sigma^2 = \mathbf{E}\xi_i^2 < \infty, S_n = \xi_1 + \cdots + \xi_n, |\xi_i| \leqslant C (\mathbf{P} - \text{a.c.}), i \leqslant n$. 证明, 对于任意 $0 \leqslant x \leqslant 2C^{-1}$, 有

$$\mathbf{E}\mathrm{e}^{xS_n} \leqslant \exp\{2^{-1}nx^2\sigma^2(1 + xC)\}.$$

在以上的条件下证明, 如果 (a_n) 是实数序列, 满足 $a_n/\sqrt{n} \to \infty$ 而 $a_n = o(n)$, 则对于任意 $\varepsilon > 0$ 和充分大的 n, 有

$$\mathbf{P}\{S_n > a_n\} > \exp\left\{-\frac{a_n^2}{2n\sigma^2}(1 + \varepsilon)\right\}.$$

8. 设 $\xi_1, \cdots, \xi_n$ 是独立同分布随机变量, 且 $\mathbf{E}\xi_i = 0, |\xi_i| \leqslant C(\mathbf{P}-\text{a.c.}), i \leqslant n; D_n = \sum_{i=1}^{n} \mathbf{D}\xi_i$. 证明, 对于 $S_n = \xi_1 + \cdots + \xi_n$, 普罗霍罗夫 (Ю. В. Прохоров) 不等式:

$$\mathbf{P}\{S_n \geqslant a_n\} \leqslant \exp\left\{-\frac{a}{2c}\arcsin\frac{ac}{2D_n}\right\}, \quad a \in \mathbb{R},$$

成立.

§5. 强大数定律的收敛速度和大偏差概率

1. 事件的频率向概率的收敛速度问题 现在考虑第一章 §6 中曾经讨论过的伯努利概型. 对于这种概型, 棣莫弗 - 拉普拉斯定理给出了标准 (正规) 偏差 $|S_n - np| \geqslant \varepsilon\sqrt{n}$ 的渐近式, 其中标准偏差即 S_n 对中心值 np 的偏差量级为 $\sqrt{n}$. 同是在第一章 §6 中, 还对所谓大偏差 $|S_n - np| \geqslant \varepsilon n$ 的概率作出了估计, 即 S_n 对 np 的 n 级偏差:

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right\} \leqslant 2\mathrm{e}^{-2n\varepsilon^2} \tag{1}$$

(见第一章 §6 的 (42) 式). 由此当然可以得到不等式

$$\mathbf{P}\left\{\sup_{m\geqslant n}\left|\frac{S_m}{m} - p\right| \geqslant \varepsilon\right\} \leqslant \sum_{m\geqslant n}\mathbf{P}\left\{\left|\frac{S_m}{m} - p\right| \geqslant \varepsilon\right\} \leqslant \frac{2}{1-\mathrm{e}^{-2\varepsilon^2}}\mathrm{e}^{-2n\varepsilon^2}, \tag{2}$$

不等式 (2) 给出了关于 "随机变量 S_n/n 以概率 1 收敛于 p 的速度" 的一定印象.

现在对于略微一般的情形, 即对于 $S_n = \xi_1 + \cdots + \xi_n$ 是独立同分布随机变量之和的情形, 讨论形如 (1), (2) 的公式正确性的问题.

2. 强大数定律中的收敛速度和大偏差概率 称随机变量 ξ 满足克拉默 (H. Cramer) 条件, 如果存在 0 的邻域, 使得对于该邻域的任何 λ, 有

$$\mathbf{E}\mathrm{e}^{\lambda\xi} < \infty \tag{3}$$

(可以证明, 这一条件等价于 "概率 $\mathbf{P}\{|\xi| > x\}$ 以指数的速度下降".)

设

$$\varphi(\lambda) = \mathbf{E}\mathrm{e}^{\lambda\xi}, \quad \psi(\lambda) = \ln\varphi(\lambda). \tag{4}$$

在集合

$$\Lambda = \{\lambda \in \mathbb{R} : \psi(\lambda) < \infty\} \tag{5}$$

的内部, 函数 $\psi(\lambda)$ 是凹 (向下凸) 函数, 且无限可微. 这时

$$\psi(0) = 0, \quad \psi'(0) = m(=\mathbf{E}\xi), \quad \psi''(\lambda) \geqslant 0.$$

建立函数

$$H(a) = \sup_{\lambda}[a\lambda - \psi(\lambda)], \quad a \in \mathbb{R}, \tag{6}$$

称做 (随机变量 ξ 的分布函数 $F = F(x)$ 的) *克拉默变换*. 函数 $H(a)$ 也是凹函数, 并且在点 $a = m$ 达到其最小值 0.

如果 $a > m$, 则

$$H(a) = \sup_{\lambda>0}[a\lambda - \psi(\lambda)].$$

因此

$$\mathbf{P}\{\xi \geqslant a\} \leqslant \inf_{\lambda>0} \mathbf{E}e^{\lambda(\xi-a)} = \inf_{\lambda>0} e^{-[a\lambda-\psi(\lambda)]} = e^{-H(a)}. \tag{7}$$

完全同样, 对于 $a < m$, 有 $H(a) = \sup\limits_{\lambda<0}[a\lambda - \psi(\lambda)]$ 和

$$\mathbf{P}\{\xi \leqslant a\} \leqslant e^{-H(a)}. \tag{8}$$

从而 (对照第一章 §6 的 (42) 式)

$$\mathbf{P}\{|\xi - m| \geqslant \varepsilon\} \leqslant 2e^{-\min\{H(m-\varepsilon), H(m+\varepsilon)\}}. \tag{9}$$

设 $\xi, \xi_1, \cdots, \xi_n$ 是独立同分布、满足克拉默条件 (3) 的随机变量, $S_n = \xi_1 + \cdots + \xi_n$, $\psi_n(\lambda) = \ln \mathbf{E}\exp\left\{\lambda \dfrac{S_n}{n}\right\}$, $\psi(\lambda) = \ln \mathbf{E}e^{\lambda\xi}$,

$$H_n(a) = \sup_{\lambda}[a\lambda - \psi_n(\lambda)], \tag{10}$$

则

$$H_n(a) = nH(a) \left(= n\sup_{\lambda}[a\lambda - \psi(\lambda)]\right),$$

而不等式 (7), (8) 和 (9) 有如下形式:

$$\mathbf{P}\left\{\frac{S_n}{n} \geqslant a\right\} \leqslant e^{-nH(a)}, \quad a > m, \tag{11}$$

$$\mathbf{P}\left\{\frac{S_n}{n} \leqslant a\right\} \leqslant e^{-nH(a)}, \quad a < m, \tag{12}$$

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - m\right| \geqslant \varepsilon\right\} \leqslant 2e^{-\min\{H(m-\varepsilon), H(m+\varepsilon)\}\times n}. \tag{13}$$

注 1 形如

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - m\right| \geqslant \varepsilon\right\} \leqslant ae^{-bn} \tag{14}$$

的结果, 其中 $a > 0$ 和 $b > 0$, 称做关于常数 a 和 b 的 "正则" 指数收敛. 在*大偏差理论*中相应的结果常用略有不同的、比较 "粗糙" 形式表示为:

$$\overline{\lim_n} \frac{1}{n} \ln \mathbf{P}\left\{\left|\frac{S_n}{n} - m\right| \geqslant \varepsilon\right\} < 0. \tag{15}$$

这种表示形式自然出于 (14) 式, 在于说明 "指数" 收敛速度, 但是不强调常数 a 和 b 的值.

现在考虑概率

$$\mathbf{P}\left\{\sup_{k\geqslant n}\frac{S_k}{k}>a\right\},\quad \mathbf{P}\left\{\inf_{k\geqslant n}\frac{S_k}{k}<a\right\},\quad \mathbf{P}\left\{\sup_{k\geqslant n}\left|\frac{S_k}{k}-m\right|>\varepsilon\right\}$$

自上侧估计的问题. 这些概率可以给出关于强大数定律中收敛速度的一定印象.

假设独立同分布非退化随机变量 $\xi,\xi_1,\xi_2,\cdots$ 满足克拉默条件 (3).

固定 $n\geqslant 1$, 设

$$\kappa=\inf\left\{k\geqslant n:\frac{S_k}{k}>a\right\},$$

当 $S_k/k\leqslant a, k\geqslant n$ 时, 设 $\kappa=\infty$.

其次, 设 a 和 $\lambda>0$, 满足

$$\lambda a-\ln\varphi(\lambda)\geqslant 0. \tag{16}$$

那么,

$$\begin{aligned}
&\mathbf{P}\left\{\sup_{k\geqslant n}\frac{S_k}{k}>a\right\}=\mathbf{P}\left(\bigcup_{k\geqslant n}\left\{\frac{S_k}{k}>a\right\}\right)=\mathbf{P}\left\{\frac{S_\kappa}{\kappa}>a,\kappa<\infty\right\}\\
&=\mathbf{P}\left\{\mathrm{e}^{\lambda S_\kappa}>\mathrm{e}^{\lambda a\kappa},\kappa<\infty\right\}=\mathbf{P}\left\{\mathrm{e}^{\lambda S_k-\kappa\ln\varphi(\lambda)}>\mathrm{e}^{\kappa[\lambda a-\ln\varphi(\lambda)]},\kappa<\infty\right\}\\
&\leqslant\mathbf{P}\left\{\mathrm{e}^{\lambda S_\kappa-\kappa\ln\varphi(\lambda)}>\mathrm{e}^{n[\lambda a-\ln\varphi(\lambda)]},\kappa<\infty\right\}\\
&\leqslant\mathbf{P}\left\{\sup_{k\geqslant n}\mathrm{e}^{\lambda S_k-k\ln\varphi(\lambda)}\geqslant\mathrm{e}^{n[\lambda a-\ln\varphi(\lambda)]}\right\}.
\end{aligned}\tag{17}$$

注意, 在证明的最后一步, 随机变量序列

$$\mathrm{e}^{\lambda S_k-k\ln\varphi(\lambda)},\quad k\geqslant 1,$$

关于 $\sigma-$ 代数流 $\mathscr{F}_k=\sigma(\xi_1,\cdots,\xi_k), k\geqslant 1$, 是鞅. (详见第七章, 特别见 §1 例 2.)

那么, 由第七章 §3 不等式 (8), 可以导出

$$\mathbf{P}\left\{\sup_{k\geqslant n}\mathrm{e}^{\lambda S_k-k\ln\varphi(\lambda)}\geqslant\mathrm{e}^{n[\lambda a-\ln\varphi(\lambda)]}\right\}\leqslant\mathrm{e}^{-n[\lambda a-\ln\varphi(\lambda)]},$$

从而, (在 (16) 式的条件下) 得不等式:

$$\mathbf{P}\left\{\sup_{k\geqslant n}\frac{S_k}{k}>a\right\}\leqslant\mathrm{e}^{-n[\lambda a-\ln\varphi(\lambda)]}. \tag{18}$$

设 $a>m$. 由于对函数 $f(\lambda)=\lambda a-\ln\varphi(\lambda)$, 有 $f(0)=0, f'(0)>0$, 则存在 $\lambda>0$ 使 (16) 式成立. 从而, 由 (18) 式可见: 如果 $a>m$, 则

$$\mathbf{P}\left\{\sup_{k\geqslant n}\frac{S_k}{k}>a\right\}\leqslant\mathrm{e}^{-n\sup\limits_{\lambda>0}[\lambda a-\ln\varphi(\lambda)]}=\mathrm{e}^{-nH(a)}. \tag{19}$$

同样, 如果 $a < m$, 则

$$\mathbf{P}\left\{\inf_{k\geqslant n}\frac{S_k}{k} < a\right\} \leqslant \mathrm{e}^{-n\sup\limits_{\lambda<0}[\lambda a-\ln\varphi(\lambda)]} = \mathrm{e}^{-nH(a)}. \tag{20}$$

由 (19) 和 (20) 式, 可见

$$\mathbf{P}\left\{\sup_{k\geqslant n}\left|\frac{S_k}{k}-m\right| > \varepsilon\right\} \leqslant 2\mathrm{e}^{-\min\{H(m-\varepsilon),H(m+\varepsilon)\}\times n}. \tag{21}$$

注 2 由 (11) 和 (19) 式右侧的相同, 使我们想到这并非偶然. 事实上, 其解释在于, 对于任何 $n \leqslant N$, 序列 $(S_k/k)_{n\leqslant k\leqslant N}$ 是可逆鞅 (见第七章 §1 练习题 5, 和第一章 §11 例 4).

3. 练习题

1. 证明不等式 (8) 和 (20).

2. 验证在集合 Λ (见 (5) 式) 的内部, 函数 $\psi(\lambda)$ 是凹 (向下凸) 函数 (若随机变量 ξ 是非退化的, 而且是严格凹函数), 并且无限可微.

3. 在随机变量 ξ 是非退化的假设条件下, 函数 $H(a)$ 在全直线上可微, 并且是凹函数.

4. 证明如下克拉姆变换的反演公式:

$$\psi(\lambda) = \sup_a[a\lambda - H(a)]$$

(对于一切 λ, 只有集合 $\Lambda = \{\lambda : \psi(\lambda) < \infty\}$ 的有限个点可能除外).

5. 设 $S_n = \xi_1 + \cdots + \xi_n$, 其中 $\xi_1, \cdots, \xi_n, n \geqslant 1$, 是独立同分布的简单随机变量, 且 $\mathbf{E}\xi_1 < 0, \mathbf{P}\{\xi_1 > 0\} > 0$. 假设 $\varphi(\lambda) = \mathbf{E}\mathrm{e}^{\lambda\xi_1}$, 而 $\inf_\lambda \varphi(\lambda) = \rho (0 < \rho < 1)$.

证明下面的 (切尔诺夫 [Чернов]) 定理:

$$\lim_n \frac{1}{n}\ln\mathbf{P}\{S_n \geqslant 0\} = \ln\rho. \tag{22}$$

6. 利用 (22) 式的结果, 证明, 对于伯努利概型 $(\mathbf{P}\{\xi_1 = 1\} = p, \mathbf{P}\{\xi_1 = 0\} = q)$, 当 $p < x < 1$ 时,

$$\lim_n \frac{1}{n}\ln\mathbf{P}\{S_n \geqslant nx\} = -H(x), \tag{23}$$

其中 (参照第一章 §6 的记号)

$$H(x) = x\ln\frac{x}{p} + (1-x)\ln\frac{1-x}{1-p}.$$

7. 设 $S_n = \xi_1 + \cdots + \xi_n, n \geqslant 1$, 其中 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量, 且 $\mathbf{E}\xi_1 = 0, \mathbf{D}\xi_1 = 1$. 设数列 $(x_n)_{n\geqslant 1}$ 满足: $x_n \to \infty$, 而当 $n \to \infty$ 时 $x_n/\sqrt{n} \to 0$.

证明

$$\mathbf{P}\{S_n \geqslant x_n\sqrt{n}\} = \mathrm{e}^{-\frac{x_n^2}{2}}(1 + y_n),$$

其中当 $n \to \infty$ 时 $y_n \to 0$.

8. 由 (23) 式证明, 对于伯努利概型: $\mathbf{P}\{\xi_1 = 1\} = p, \mathbf{P}\{\xi_1 = 0\} = q$,

(a) 当 $p < x < 1$, 且 $x_n = n(x - p)$ 时,

$$\mathbf{P}\{S_n \geqslant np + x_n\} = \exp\left\{-nH\left(p + \frac{x_n}{n}\right)(1 + o(1))\right\}; \tag{24}$$

(b) 当 $x_n = a_n\sqrt{npq}$ 且 $a_n \to \infty, a_n/\sqrt{n} \to 0$ 时,

$$\mathbf{P}\{S_n \geqslant np + x_n\} = \exp\left\{-\frac{x_n^2}{2npq}(1 + o(1))\right\}. \tag{25}$$

将 (24) 和 (25) 两式与第一章 §6 中相应的结果进行比较.



第五章　强 (狭义) 平稳随机序列和遍历理论



在概率论的框架之外, 可以将强平稳随机序列理论, 表述为可测测度空间上的单参数保测变换群理论, 它与动态系统的一般理论和遍历性理论临近.

《数学百科全书》, (中译本) 第 4 卷, 第 985 页 [121]

---

§1. 强 (狭义) 平稳随机序列. 保测变换

1. 强平稳随机变量序列 设 $(\Omega,\mathscr{F},\mathbf{P})$ 是一概率空间, $\xi=(\xi_1,\xi_2,\cdots)$ 是一随机变量序列, 或简称*随机序列*. 以 $\theta_k\xi$ 表示序列 $(\xi_{k+1},\xi_{k+2},\cdots)$.

定义 1 随机变量序列 ξ 称做*强平稳的*, 如果对于任意 $k\geqslant 1$, $\theta_k\xi$ 和 ξ 有相同的概率分布:

$$\mathbf{P}\{(\xi_1,\xi_2,\cdots)\in B\}=\mathbf{P}\{(\xi_{k+1},\xi_{k+2},\cdots)\in B\},\quad B\in\mathscr{B}(\mathbb{R}^\infty).$$

由独立同分布随机变量 $\xi=(\xi_1,\xi_2,\cdots)$ 形成的序列, 是平稳序列 ξ 最简单的例子. 从这样的序列出发, 可以构造广泛的平稳序列类 $\eta=(\eta_1,\eta_2,\cdots)$, 其中 $\eta_k=g(\xi_k,\xi_{k+1},\cdots,\xi_{k+n})$, 而 $g(x_1,\cdots,x_n)$ 是任意博雷尔函数.

如果 $\xi=(\xi_1,\xi_2,\cdots)$ 是独立同分布随机变量序列, $\mathbf{E}|\xi_1|<\infty,\mathbf{E}\xi_1=m$, 则根据强大数定律, 以概率 1

$$\frac{\xi_1+\cdots+\xi_n}{n}\to m,\quad n\to\infty.$$

伯克霍夫 (G. D. Birkhoff) 在 1931 年作为*力学定理*, 得到了这一结果出色的推广. 所谓力学定理, 涉及动态系统的 "相对到达时间" 的性质, 这里的动态系统是由微分方程描述的, 而假设微分方程有积分变式 ("守恒系统").

辛钦于 1932 年紧接着证明了, 实际上伯克霍夫定理可以推广到更一般情形: "多维空间在自身中的运动, 并且保持集合的测度不变".

我们将既在 "动态系统" 理论的框架内, 又在 "强平稳随机序列" 的框架内, 同时叙述伯克霍夫和辛钦的结果.

这时基本着重点放在该理论有关 "遍历性" 的结果上.

2. 保测变换 设 $(\Omega,\mathscr{F},\mathbf{P})$ 是一 (完备) 概率空间.

定义 2 如果对于任何 $A\in\mathscr{F}$, 有

$$T^{-1}A=\{\omega:T\omega\in A\}\in\mathscr{F},$$

则空间 Ω 到自身的影射 T 称做*可测的*.

定义 3 如果对于任何 $A \in \mathscr{F}$, 有

$$\mathbf{P}(T^{-1}A) = \mathbf{P}(A),$$

则可测影射 T 称做保测变换 (morphism).

设 T 是保测变换, T^n 是其 n 次幂, 而 $\xi_1 = \xi_1(\omega)$ 是一随机变量. 设 $\xi_n = \xi_1(T^{n-1}\omega), n \geqslant 2$, 考虑序列 $\xi = (\xi_1, \xi_2, \cdots)$. 我们证明此序列是平稳的.

事实上, 设 $A = \{\omega : \xi \in B\}, A_1 = \{\omega : \theta_1\xi \in B\}$, 其中 $B \in \mathscr{B}(\mathbb{R}^\infty)$, 则 $\omega \in A_1$ 当且仅当 $T\omega \in A$, 即 $A_1 = T^{-1}A$. 由于 $\mathbf{P}(T^{-1}A) = \mathbf{P}(A)$, 可见 $\mathbf{P}(A_1) = \mathbf{P}(A)$. 同样, 对于任意事件 $A_k = \{\omega : \theta_k\xi \in B\}, k \geqslant 2$, 有 $\mathbf{P}(A_k) = \mathbf{P}(A)$.

这样, 所引进的保测变换, 为建立强平稳序列提供了可能性.

在一定意义上有相反的结果: 对每一平稳序列 ξ, 对所考虑概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$, 存在新概率空间 $(\widetilde{\Omega}, \widetilde{\mathscr{F}}, \widetilde{\mathbf{P}})$, 以及随机变量 $\widetilde{\xi}_1(\widetilde{\omega})$ 和保测变换 $\widetilde{T}$, 使随机序列 $\widetilde{\xi} = \{\widetilde{\xi}_1(\widetilde{\omega}), \widetilde{\xi}_1(\widetilde{T}\widetilde{\omega}), \cdots\}$ 与随机序列 $\xi = \{\xi_1(\omega), \xi_2(\omega), \cdots\}$ 有相同的分布.

事实上, 作为 $\widetilde{\Omega}$ 取 "坐标" 空间 $\mathbb{R}^\infty$, 并设 $\widetilde{\mathscr{F}}(\mathbb{R}^\infty), \widetilde{\mathbf{P}} = P_\xi$, 其中 $P_\xi(B) = \mathbf{P}\{\omega : \xi \in B\}, B \in \mathscr{B}(\mathbb{R}^\infty)$. 空间 $\widetilde{\Omega}$ 的变换 $\widetilde{T}$, 决定于关系式 $\widetilde{T}(x_1, x_2, \cdots) = (x_2, x_3, \cdots)$. 对于 $\widetilde{\omega} = (x_1, x_2, \cdots)$, 记

$$\widetilde{\xi}_1(\widetilde{\omega}) = x_1, \quad \widetilde{\xi}_n(\widetilde{\omega}) = \widetilde{\xi}_1(\widetilde{T}^{n-1}\widetilde{\omega}), \quad n \geqslant 2.$$

现在设 $A = \{\widetilde{\omega} : (x_1, \cdots, x_k) \in B\}, B \in \mathscr{B}(\mathbb{R}^k)$, 且 $\widetilde{T}^{-1}A = \{\widetilde{\omega} : (x_2, \cdots, x_{k+1}) \in B\}$. 那么, 由平稳性, 可见

$$\widetilde{\mathbf{P}}(A) = \mathbf{P}\{\omega : (\xi_1, \cdots, \xi_k) \in B\} = \mathbf{P}\{\omega : (\xi_2, \cdots, \xi_{k+1}) \in B\} = \mathbf{P}(\widetilde{T}^{-1}A),$$

即 $\widetilde{T}$ 是保测变换. 由于对于任意 k,

$$\widetilde{\mathbf{P}}\{\widetilde{\omega} : (\widetilde{\xi}_1, \cdots, \widetilde{\xi}_k) \in B\} = \mathbf{P}\{\omega : (\xi_1, \cdots, \xi_k) \in B\},$$

则由此可见 ξ 和 $\widetilde{\xi}$ 有相同的分布.

下面是保测变换的例子.

例 1 设 $\Omega = \{\omega_1, \cdots, \omega_n\}, n \geqslant 2$, 是由有限个点构成的集合, $\mathscr{F}$ 是 Ω 中一切子集的 $\sigma-$ 代数, $T\omega_i = \omega_{i+1}, 1 \leqslant i \leqslant n-1$, 而 $T\omega_n = \omega_1$. 如果 $\mathbf{P}(\omega_i) = 1/n$, 则 T 是保测变换.

例 2 设 $\Omega = [0,1), \mathscr{F} = \mathscr{B}([0,1)), \mathbf{P}$ 是勒贝格测度, $\lambda \in [0,1)$, 则 $Tx = (x + \lambda)$ 是保测变换.

现在考虑导致研究保测变换的物理前提条件.

假设某一系统按照一定运动规律 (在离散时间) 演变, 并设想 Ω 是该系统的状态 ω 的相空间. 那么, 如果 ω 是系统在时刻 $n = 1$ 的状态, 则 $T^n\omega$ 是经过 n 步系统进入

的状态, 其中 T 是 (该运动规律诱导的) 推移算子. 其次, 假如 A 是某一 "状态 ω 的集合", 则 $T^{-1}A=\{\omega: T\omega\in A\}$ 根据其定义是经一步到达集合 A 的一切 "初始" 状态 ω 的集合. 因此, 假如把 Ω 视为 "不可压缩的液体", 则条件 $\mathbf{P}(T^{-1}A)=\mathbf{P}(A)$ 可以视为完全自然的 "体积" 保持不变的条件. (对于经典的封闭哈密顿 (W. R. Hamilton) 系统, 著名的刘维尔 (J. Liouville) 定理断定, 相应的变换 T 是保持勒贝格测度不变的变换.)

3. 关于保测变换的庞加莱定理　下面关于 "常返性" 的庞加莱 (J. H. Poincaré) 定理 (1912), 是有关保测变换最早的成果之一.

定理　设 $(\Omega,\mathscr{F},\mathbf{P})$ 是一概率空间, T 是保测变换, $A\in\mathscr{F}$. 那么, 对于无限多个 $n\geqslant 1$ 和几乎一切点 $\omega\in A$, 有 $T^n\omega\in A$.

证明　记 $C=\{\omega\in A: T^n\omega\notin A$, 对于一切 $n\geqslant 1\}$. 由于对于任意 $n\geqslant 1, C\bigcap T^{-n}C=\varnothing$, 则 $T^{-m}C\bigcap T^{-(m+n)}C=T^{-m}(C\bigcap T^{-n}C)=\varnothing$. 这样, 序列 $\{T^{-n}C\}$ 由有相同 $\mathbf{P}-$ 测度的不相交集合构成. 因此,

$$\sum_{n=1}^{\infty}\mathbf{P}(C)=\sum_{n=1}^{\infty}\mathbf{P}(T^{-n}C)\leqslant\mathbf{P}(\Omega)=1,$$

从而 $\mathbf{P}(C)=0$. 从而, 对于几乎一切点 $\omega\in A$ 和至少对于一个 $n\geqslant 1, T^n\omega\in A$. 由此可见, 对于无限个 $n\geqslant 1$, 有 $T^n\omega=A$.

将上面得到的结果用于变换 $T^k\omega, k\geqslant 1$. 那么, 对于每一个点 $\omega\in A\backslash N$, 其中 N 是 0 概率集合 (并且 N 是对应于不同 k 的相应集合的并), 存在这样的 n_k, 使 $(T^k)^{n_k}\omega\in A$. 由此显然可以得到, 对于无限多个 n, 有 $T^n\omega\in A$. □

系　设 $\xi(\omega)\geqslant 0$, 则在集合 $\{\omega:\xi(\omega)>0\}$ 上

$$\sum_{k=0}^{\infty}\xi(T^k\omega)=\infty\quad(\mathbf{P}-\text{a.c.}).$$

实际上, 设 $A_n=\{\omega:\xi(\omega)\geqslant 1/n\}$, 那么, 根据上面的定理在集合 A_n 上, 有

$$\sum_{k=0}^{n}\xi(T^k\omega)=\infty\quad(\mathbf{P}-\text{a.c.}),$$

如果令当 $n\to\infty$ 时, 得所要证明的结果.

注　假如将概率测度 $\mathbf{P}$ 换成任意有限测度 $\mu(\mu(\Omega)<\infty)$, 则定理仍然成立.

4. 练习题

1. 设 T 是保测变换, $\xi=\xi(\omega)$ 是一随机变量, 并且有数学期望 $\mathbf{E}\xi(\omega)$. 证明 $\mathbf{E}\xi(\omega)=\mathbf{E}\xi(T\omega)$.

2. 证明, 例 1 和例 2 中变换 T 是保测变换

3. 设 $\Omega=[0,1),\mathscr{F}=\mathscr{B}([0,1)),\mathbf{P}$ 是具有连续分布函数测度, 证明变换

$$Tx=\lambda x(0<\lambda<1)\quad 和\quad Tx=x^2$$

都不是保测变换.

4. 设 Ω 是一切实数序列 $\omega=(\cdots,\omega_{-1},\omega_0,\omega_1,\cdots)$ 的集合, $\mathscr{F}$ 是由可测柱集诱导的 $\sigma-$ 代数, 其中的可测柱集为 $\{\omega:(\omega_k,\cdots,\omega_{k+n-1})\in B_n\}(n=1,2,\cdots;k=0,\pm1,\pm2,\cdots)$, 而 $B_n\in\mathscr{B}(\mathbb{R}^n)$. 以 $\mathbf{P}$ 表示 $(\Omega,\mathscr{F})$ 上的概率测度, 而双侧变换 T 决定于公式:

$$T(\cdots,\omega_{-1},\omega_0,\omega_1,\cdots)=(\cdots,\omega_0,\omega_1,\omega_2,\cdots).$$

证明, T 是保测变换, 当且仅当对于一切 $n=1,2,\cdots,k=0,\pm1,\pm2,\cdots$ 和 $B_n\in\mathscr{B}(\mathbb{R}^n)$

$$\mathbf{P}\{\omega:(\omega_0,\cdots,\omega_{n-1})\in B_n\}=\mathbf{P}\{\omega:(\omega_k,\cdots,\omega_{k+n-1})\in B_n\}.$$

5. 设 $\xi_0,\xi_1,\cdots$ 是一平稳随机元序列, $\xi_n(n\geqslant 0)$ 的值属于博雷尔空间 S (见第二章 §7 定义 9). 证明, 可以构造 (有可能在原概率空间的扩展空间上) 在 S 中取值的随机元素 $\xi_{-1},\xi_{-2},\cdots$, 使双侧序列 $\cdots,\xi_{-1},\xi_0,\xi_1,\cdots$ 是平稳的.

6. 设 T 是 $(\Omega,\mathscr{F},\mathbf{P})$ 上的可测变换, $\mathscr{E}$ 是由 $\mathscr{F}(\pi(\mathscr{E})=\mathscr{F})$ 诱导的 Ω 子集的 $\pi-$ 系. 证明, 如果对于 $A\in\mathscr{E}$ 等式 $\mathbf{P}(T^{-1}A)=\mathbf{P}(A)$ 成立, 则此等式对于 $A\in\mathscr{F}(=\pi(\mathscr{E}))$ 也成立.

7. 设 T 是在 $(\Omega,\mathscr{F},\mathbf{P})$ 上的可测变换, $\mathscr{G}$ 是 $\mathscr{F}$ 的子 $\sigma-$ 代数. 证明对每个 $A\in\mathscr{F}$, 有

$$\mathbf{P}(A|\mathscr{G})(T\omega)=\mathbf{P}(T^{-1}A|T^{-1}\mathscr{G})(\omega)\quad(\mathbf{P}-\text{a.c.}).\tag{1}$$

特别, 设 $\Omega=\mathbb{R}^\infty$ 是数列 $\omega=(\omega_0,\omega_1,\cdots)$ 的空间, 而 $\xi_k(\omega)=\omega_k$ 的; T是推移变换: $T(\omega_0,\omega_1,\cdots)=(\omega_1,\omega_2,\cdots)$, 换句话说, 若 $\xi_k(\omega)=\omega_k$, 则 $\xi_k(T\omega)=\omega_{k+1}$. 那么, 等式 (1) 有如下形式

$$\mathbf{P}(A|\xi_n)(T\omega)=\mathbf{P}(T^{-1}A|\xi_{n+1})(\omega)\quad(\mathbf{P}-\text{a.c.}).$$

8. 设 T 是在 $(\Omega,\mathscr{F})$ 上的可测变换, $\mathscr{P}$ 是所有概率测度 $\mathbf{P}$ 的集合, 而关于概率测度 $\mathbf{P}$, 变换 T 是保持测度 $\mathbf{P}$ 不变的变换. 证明:

(a) 集合 $\mathscr{P}$ 是凸的;

(b) 变换 T 关于概率测度 $\mathbf{P}$ 是遍历变换, 当且仅当 $\mathbf{P}$ 是集合 $\mathscr{P}$ 的边界点 (即不能表示为如下形式的点: $\mathbf{P}=\lambda_1\mathbf{P}_1+\lambda_2\mathbf{P}_2,\lambda_1>0,\lambda_2>0,\lambda_1+\lambda_2=1,\mathbf{P}_1\neq\mathbf{P}_2,\mathbf{P}_1,\mathbf{P}_2\in\mathscr{P}$).

§2. 遍历性与混合性

1. 保测变换的遍历性 在整个这一节中, 我们都将以 T 表示作用于概率空间 $(\Omega,\mathscr{F},\mathbf{P})$ 上的保测变换.

定义 1 集合 $A\in\mathscr{F}$ 称做不变的, 如果 $T^{-1}A=A$. 集合 $A\in\mathscr{F}$ 称做几乎不变的, 如果 A 和 $T^{-1}A=A$ 只相差一 0 测集, 即 $\mathbf{P}(A\triangle T^{-1}A)=0$.

不难验证, 不变 (几乎不变) 集合类 $\mathscr{S}$ (相应地 $\mathscr{S}^*$) 构成 $\sigma-$ 代数.

定义 2 保测变换 T 称做遍历的 (或度量可递变换), 如果每一个不变集合 A 的测度只能为 0 或 1.

定义 3 随机变量 $\eta = \eta(\omega)$ 称做不变的 (或几乎不变的), 如果对于一切 $\omega \in \Omega$ (对于几乎一切 $\omega \in \Omega$)$\eta(\omega) = \eta(T\omega)$.

下面的引理建立不变集合与几乎不变集合之间的联系.

引理 1 如果 A 是几乎不变集合, 则存在一几乎不变集合 B, 使 $\mathbf{P}(A\triangle B) = 0$.

证明 设 $B = \overline{\lim} T^{-n}A$, 则 $T^{-1}B = \overline{\lim} T^{-(n+1)}A = B$, 即 $B \in \mathscr{S}$. 不难验证

$$A\triangle B \subseteq \bigcup_{k=0}^{\infty}(T^{-k}A\triangle T^{-(k+1)}A).$$

由于 $\mathbf{P}(T^{-k}A\triangle T^{-(k+1)}A) = \mathbf{P}(A\triangle T^{-1}A) = 0$, 可见 $\mathbf{P}(A\triangle B) = 0$. □

引理 2 变换 T 是遍历的, 当且仅当每一个几乎不变集合 A 的测度只能为 0 或 1.

证明 设 $A \in \mathscr{S}^*$, 则根据引理 1, 存在一几乎不变集合 B, 使 $\mathbf{P}(A\triangle B) = 0$. 但是, 由于变换 T 是遍历的, 故 $\mathbf{P}(B) = 0$ 或 1. 由于 $\mathscr{S} \subseteq \mathscr{S}^*$, 故逆命题显然成立. □

定理 1 设 T 是保测变换. 下列各个条件等价:

(1) T 是遍历的;

(2) 每一个几乎不变随机变量以概率 $\mathbf{P}$ 为 1 是常数;

(3) 每一个不变随机变量以概率 $\mathbf{P}$ 为 1 是常数.

证明 (1)⇒(2). 设 T 是遍历的, 而 ξ 是几乎不变的, 即以概率 1 有 $\xi(\omega) = \xi(T\omega)$. 那么, 对于任何 $c \in \mathbb{R}$, 集合 $A_c = \{\omega : \xi(\omega) \leqslant c\} \in \mathscr{S}^*$, 而根据引理 2, $\mathbf{P}(A_c) = 0$ 或 1. 设 $C = \sup\{c : \mathbf{P}(A_c) = 0\}$. 当 $c \uparrow \infty$ 时 $A_c \uparrow \Omega$, 而当 $c \downarrow -\infty$ 时 $A_c \downarrow \varnothing$, 故 $C < \infty$.

那么,

$$\mathbf{P}\{\omega : \xi(\omega) < C\} = \mathbf{P}\left(\bigcup_{n=1}^{\infty}\left\{\xi(\omega) \leqslant C - \frac{1}{n}\right\}\right) = 0;$$

类似可得, $\mathbf{P}\{\omega : \xi(\omega) > C\} = 0$. 于是, $\mathbf{P}\{\omega : \xi(\omega) = C\} = 1$.

(2)⇒(3). 显然.

(3)⇒(1). 设 $A \in \mathscr{S}$, 则 I_A 是不变随机变量, 因此以概率 1 有 $I_A = 0$ 或 $I_A = 1$, 故 $\mathbf{P}(A) = 0$ 或 1. □

注 如果在定理中的随机变量有界, 则定理的结论仍然成立.

为演示该定理的应用, 我们看下面的例子.

例 设 $\Omega = [0,1), \mathscr{F} = \mathscr{B}([0,1)), \mathbf{P}$ 是勒贝格测度, 而 $T\omega = (\omega + 1) \bmod 1$. 证明 T 是遍历变换, 当且仅当 λ 是无理数.

设 $\xi=\xi(\omega)$ 是不变随机变量, $\mathbf{E}\xi^2(\omega)<\infty$. 熟知, 具有的 $\mathbf{E}\xi^2(\omega)<\infty$ 函数 $\xi(\omega)$ 的傅里叶级数 $\sum\limits_{n=-\infty}^{\infty} c_n \mathrm{e}^{2\pi i n\omega}$ 均方收敛, 且 $\sum|c_n|^2<\infty$. 由于 T 是保测变换 (§1 例 2), 则由于所假设的随机变量 $\xi=\xi(\omega)$ 的不变性 (§1 练习题 1). 可见

$$\begin{aligned} c_n &= \mathbf{E}\xi(\omega)\mathrm{e}^{-2\pi i n\omega} = \mathbf{E}\xi(T\omega)\mathrm{e}^{-2\pi i nT\omega} = \mathrm{e}^{-2\pi i n\lambda}\mathbf{E}\xi(T\omega)\mathrm{e}^{-2\pi i n\omega} \\ &= \mathrm{e}^{-2\pi i n\lambda}\mathbf{E}\xi(\omega)\mathrm{e}^{-2\pi i n\omega} = c_n\mathrm{e}^{-2\pi i n\lambda}. \end{aligned}$$

因此 $c_n(1-\mathrm{e}^{-2\pi i n\lambda})=0$. 根据条件 λ 是无理数, 即对于一切 $n\neq 1, \mathrm{e}^{-2\pi i n\lambda}\neq 1$. 于是, $c_n=0, n\neq 1, \xi(\omega)=c_0(\mathbf{P}-\text{a.c.})$, 而根据定理 1, T 是遍历变换.

另一方面, 假如 λ 是有理数, 即 $\lambda=k/m$, 其中 k 和 m 是整数. 考虑集合

$$A=\bigcup_{k=0}^{2m-2}\left\{\omega:\frac{k}{2m}\leqslant\omega<\frac{k+1}{2m}\right\}.$$

显然, 集合 A 不变的, 然而 $\mathbf{P}(A)=1/2$. 从而, T 不是遍历的.

2. 变换的混合性及其与遍历性的关系

定义 4 称保测变换 T 为混合 (具有混合性的), 如果对于任意 $A,B\in\mathscr{F}$, 有

$$\lim_{n\to\infty}\mathbf{P}(A\bigcap T^{-n}B)=\mathbf{P}(A)\mathbf{P}(B). \tag{1}$$

下面的定理指出了变换的混合性与遍历性的关系.

定理 2 *具有混合性的任何变换 T 都是遍历的.*

证明 设 $A\in\mathscr{F}, B\in\mathscr{S}$, 则 $B=T^{-n}B, n\geqslant 1$, 即对于一切 $n\geqslant 1$, 有

$$\mathbf{P}(A\bigcap T^{-n}B)=\mathbf{P}(A\bigcap B).$$

因此由于 (1) 式, 当 $A=B$ 时 $\mathbf{P}(B)=\mathbf{P}^2(B)$, 于是, $\mathbf{P}(B)=0$ 或 1. □

3. 练习题

1. 证明随机变量 $\xi=\xi(\omega)$ 是不变的, 当且仅当它 $\mathscr{S}-$ 可测.

2. 证明集合 A 是几乎不变的, 当且仅当 $\mathbf{P}(T^{-1}A\backslash A)=0$.

3. 证明变换 T 具有混合性, 当且仅当对于任何两个随机变量 ξ 和 η : $\mathbf{E}\xi^2<\infty, \mathbf{E}\eta^2<\infty$, 当 $n\to\infty$ 时, 有

$$\mathbf{E}\xi(T^{-1}\omega)\eta(\omega)\to\mathbf{E}\xi(\omega)\mathbf{E}\eta(\omega).$$

4. 举例说明, 保测遍历变换未必具有混合性质.

5. 设 T 是空间 $(\Omega,\mathscr{F},\mathbf{P})$ 上的可测变换, $\mathscr{A}$ 是 Ω 的子集的 $\sigma-$ 代数, 而 $\sigma(\mathscr{A})=\mathscr{F}$. 假如定义 4 仅要求性质

$$\lim_{n\to\infty}\mathbf{P}(A\bigcap T^{-n}B)=\mathbf{P}(A)\mathbf{P}(B)$$

对于 $A, B \in \mathscr{A}$ 成立. 证明此性质对于一切 $A, B \in \mathscr{F} = \sigma(\mathscr{A})$ 成立. (从而变换 T 是混合).

证明如果 $\mathscr{A}$ 是 $\pi-$ 系且 $\pi(\mathscr{A}) = \mathscr{F}$, 则上述结论仍然成立.

6. 设 A 是几乎不变集合. 证明, 当且仅当对于一切 $n = 1, 2, \cdots, T^n\omega \in A$, 以概率 1, 有 $\omega \in A$. (对照 §1 的定理.)

7. 举一空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 上的保测变换 T 的例, 使得:

(a) 由 $A \in \mathscr{F}$ 绝对得不出 $TA \in \mathscr{F}$;

(b) 由 $A \in \mathscr{F}$ 和 $TA \in \mathscr{F}$ 绝对得不出 $\mathbf{P}(A) = \mathbf{P}(TA)$.

§3. 遍历性定理

1. 毕达哥拉斯 – 辛钦定理

定理 1 (毕达哥拉斯 [Pythagoras] – 辛钦) 设 T 保测变换, 而 $\xi = \xi(\omega)$ 是随机变量, $\mathbf{E}|\xi| < \infty$. 那么,

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} \xi(T^k\omega) = \mathbf{E}(\xi|\mathscr{S}) \quad (\mathbf{P} - \text{a.c.}). \tag{1}$$

而且, 若 T 是遍历的, 则

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} \xi(T^k\omega) = \mathbf{E}\xi \quad (\mathbf{P} - \text{a.c.}). \tag{2}$$

该定理的证明本质上基于下面的引理, 其简单的证明是加尔西亚 (A. M. Garsia, 1965 年) 得到的.

引理 (最大遍历性定理) 设 T 保测变换, 而 $\xi = \xi(\omega)$ 是随机变量, $\mathbf{E}|\xi| < \infty$, 且

$$S_k(\omega) = \xi(\omega) + \xi(T\omega) + \cdots + \xi(T^{k-1}\omega),$$
$$M_k(\omega) = \max\{0, S_1(\omega), \cdots, S_k(\omega)\}.$$

那么, 对于任意 $n \geqslant 1$, 有

$$\mathbf{E}[\xi(\omega) I_{\{M_n>0\}}(\omega)] \geqslant 0.$$

证明 如果 $n > k$, 则 $M_n(T\omega) \geqslant S_k(T\omega)$, 从而

$$\xi(\omega) + M_n(T\omega) \geqslant \xi(\omega) + S_k(T\omega) = S_{k+1}(\omega).$$

因为显然 $\xi(\omega) \geqslant S_1(\omega) - M_n(T\omega)$, 所以

$$\xi(\omega) \geqslant \max\{S_1(\omega), \cdots, S_n(\omega)\} - M_n(T\omega).$$

注意到, 由于 $\{M_n(\omega)>0\}=\{\max[S_1(\omega),\cdots,S_n(\omega)]>0\}$, 可见

$$\begin{aligned}\mathbf{E}[\xi(\omega)I_{\{M_n>0\}}(\omega)]&\geqslant\mathbf{E}[\max\{S_1(\omega),\cdots,S_n(\omega)\}-M_n(T\omega)I_{\{M_n>0\}}(\omega)]\\&\geqslant\mathbf{E}[M_n(\omega)-M_n(T\omega)I_{\{M_n(\omega)>0\}}(\omega)]\\&\geqslant\mathbf{E}[M_n(\omega)-M_n(T\omega)]=0,\end{aligned}$$

其中用到, 如果 T 是保测变换, 则 $\mathbf{E}M_n(\omega)=\mathbf{E}M_n(T\omega)$ (§1 的练习题 1). □

证明定理 1 假设 $\mathbf{E}(\xi|\mathscr{I})=0$ (否则用 $\xi-\mathbf{E}(\xi|\mathscr{I})$ 替换 ξ). 设

$$\overline{\eta}=\varlimsup_n\frac{S_n}{n},\quad \underline{\eta}=\varliminf_n\frac{S_n}{n}.$$

为证明 (1) 式, 只需证明以概率 1, 由

$$0\leqslant\underline{\eta}\leqslant\overline{\eta}\leqslant 0.$$

考虑随机变量 $\overline{\eta}=\overline{\eta}(\omega)$. 由于 $\overline{\eta}=\eta(T\omega)$, 可见 $\overline{\eta}$ 是不变的. 从而, 对于每一个 $\varepsilon>0$, 集合 $A_\varepsilon=\{\overline{\eta}(\omega)>\varepsilon\}$ 是不变的. 现在引进新的随机变量:

$$\xi^*(\omega)=[\xi(\omega)-\varepsilon]I_{A_\varepsilon}(\omega),$$

并设

$$\begin{aligned}S_k^*(\omega)&=\xi^*(\omega)+\xi^*(T\omega)+\cdots+\xi^*(T^{k-1}\omega),\\M_k^*(\omega)&=\max\{0,S_1^*(\omega),\cdots,S_k^*(\omega)\}.\end{aligned}$$

那么, 根据引理, 对于任意 $n\geqslant1$, 有

$$\mathbf{E}[\xi^*I_{\{M_n^*>0\}}]\geqslant 0.$$

而当 $n\to\infty$ 时

$$\begin{aligned}\{M_n^*>0\}&=\left\{\sup_{1\leqslant k\leqslant n}S_k^*>0\right\}\uparrow\left\{\sup_{k\geqslant1}S_k^*>0\right\}\\&=\left\{\sup_{k\geqslant1}\frac{S_k^*}{k}>0\right\}=\left\{\sup_{k\geqslant1}\frac{S_k}{k}>\varepsilon\right\}\bigcap A_\varepsilon=A_\varepsilon,\end{aligned}$$

其中由于 $\sup\limits_{k\geqslant1}S_k/k>\overline{\eta}$, 而且 $A_\varepsilon=\{\omega:\overline{\eta}>\varepsilon\}$, 可见最后一等式成立.

此外, $\mathbf{E}|\xi^*|\leqslant\mathbf{E}|\xi|+\varepsilon$. 因此, 根据控制收敛定理, 有

$$0\leqslant\mathbf{E}[\xi^*I_{\{M_n^*>0\}}]\to\mathbf{E}[\xi^*I_{A_\varepsilon}].$$

于是,

$$\begin{aligned}0\leqslant\mathbf{E}[\xi^*I_{A_\varepsilon}]&=\mathbf{E}[(\xi-\varepsilon)I_{A_\varepsilon}]=\mathbf{E}[\xi I_{A_\varepsilon}]-\varepsilon\mathbf{P}(A_\varepsilon)\\&=\mathbf{E}[\mathbf{E}(\xi|\mathscr{I})I_{A_\varepsilon}]-\varepsilon\mathbf{P}(A_\varepsilon)=-\varepsilon\mathbf{P}(A_\varepsilon),\end{aligned}$$

因此 $\mathbf{P}(A_\varepsilon)=0$, 从而, 有 $\mathbf{P}\{\overline{\eta}\leqslant 0\}=1$.

类似地, 用 $-\xi(\omega)$ 代替 $\xi(\omega)$, 有

$$\overline{\lim_n}\left(-\frac{S_n}{n}\right)=-\underline{\lim_n}\frac{S_n}{n}=-\underline{\eta},$$

而 $\mathbf{P}\{-\underline{\eta}\leqslant 0\}=1$, 即 $\mathbf{P}\{\underline{\eta}\geqslant 0\}=1$. 因而, 以概率 1 有 $0\leqslant\underline{\eta}\leqslant\overline{\eta}\leqslant 0$. 于是, 证明了定理的第一个结论.

为证明了定理的第二个结论, 只需注意到, 由于 $\mathbf{E}(\xi|\mathscr{S})$ 是不变随机变量, 可见若 T 是遍历的, 则以概率 1 有 $\mathbf{E}(\xi|\mathscr{S})=\mathbf{E}\xi$. □

系 保测变换 T 是遍历的, 当且仅当对于任意 $A,B\in\mathscr{F}$, 有

$$\lim_n\frac{1}{n}\sum_{k=0}^{n-1}\mathbf{P}(A\bigcap T^{-k}B)=\mathbf{P}(A)\mathbf{P}(B).\tag{3}$$

为证明 T 的遍历性, 在 (3) 式中设 $A=B\in\mathscr{S}$, 则 $A\bigcap T^{-k}B=B$, 故 $\mathbf{P}(B)=\mathbf{P}^2(B)$, 因此 $\mathbf{P}(B)=0$ 或 1. 相反, 设 T 是遍历的, 则对于随机变量 $\xi=I_B(\omega)$, 由 (2) 式可见以概率 1 有

$$\lim_n\frac{1}{n}\sum_{k=0}^{n-1}I_{T^{-k}B}(\omega)=\mathbf{P}(B).$$

于该式两侧在集合上同时求积分, 并利用控制收敛定理, 得所要证明的关系式 (3).

2. 在平均收敛意义下的毕达哥拉斯 – 辛钦定理 现在证明, 在定理 1 的条件中的 (1) 式和 (2) 式, 不仅以概率 1 收敛, 而且平均收敛也成立. (下面定理 3 的证明将要用到这一结果.)

定理 2 设 T 保测变换, 而 $\xi=\xi(\omega)$ 是随机变量, $\mathbf{E}|\xi|<\infty$. 那么,

$$\mathbf{E}\left|\frac{1}{n}\sum_{k=0}^{n-1}\xi(T^k\omega)-\mathbf{E}(\xi|\mathscr{S})\right|\to 0,\quad n\to\infty.\tag{4}$$

而且, 若 T 是遍历的, 则

$$\mathbf{E}\left|\frac{1}{n}\sum_{k=0}^{n-1}\xi(T^k\omega)-\mathbf{E}(\xi)\right|\to 0,\quad n\to\infty.\tag{5}$$

证明 对于任意 $\varepsilon>0$, 存在一有界随机变量 $\eta(|\eta(\omega)|\leqslant M)$, 使 $\mathbf{E}|\xi-\eta|\leqslant\varepsilon$. 那么,

$$\mathbf{E}\left|\frac{1}{n}\sum_{k=0}^{n-1}\xi(T^k\omega)-\mathbf{E}(\xi|\mathscr{S})\right|\leqslant\mathbf{E}\left|\frac{1}{n}\sum_{k=0}^{n-1}[\xi(T^k\omega)-\eta(T^k\omega)]\right|$$
$$+\mathbf{E}\left|\frac{1}{n}\sum_{k=0}^{n-1}[\eta(T^k\omega)-\mathbf{E}(\eta|\mathscr{S})]\right|+\mathbf{E}|\mathbf{E}(\xi|\mathscr{S})-\mathbf{E}(\eta|\mathscr{S})|.\tag{6}$$

由于 $|\eta| \leqslant M$, 则由控制收敛定理和 (1) 式可见, 当 $n \to \infty$ 时, (6) 式中不等号的右侧第二项趋向 0. 至于第一项和第三项中, 每一项都小于或等于 ε. 因此, 对于充分大的 n, (6) 式小于 2ε, 从而 (4) 式得证. 最后, 如果若 T 是遍历的, 则由 (4) 式, 以及定理 1 之证明末尾的说明: "若 T 是遍历的, 则以概率 1 有 $\mathbf{E}(\xi|\mathscr{S}) = \mathbf{E}\xi$", 立即得 (5) 式. □

3. 遍历性定理 现在考虑遍历性对于强平稳随机序列正确性的问题. 设 $\xi = (\xi_1, \xi_2, \cdots)$ 是定义在概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 上的强平稳随机序列. 一般, 在 $(\Omega, \mathscr{F}, \mathbf{P})$ 上有可能并不存在保测变换, 故不能直接运用定理 1. 不过, 在 §1 中已经指出, 可以考虑建立这样的 (坐标) 概率空间 $(\widetilde{\Omega}, \widetilde{\mathscr{F}}, \widetilde{\mathbf{P}})$、随机序列 $\widetilde{\xi} = (\widetilde{\xi}_1, \widetilde{\xi}_2, \cdots)$ 和保测变换 $\widetilde{T}$, 其中 $\widetilde{\xi}_n(\widetilde{\omega}) = \widetilde{\xi}_1(\widetilde{T}^{-1}\widetilde{\omega})$, 使得 ξ 和 $\widetilde{\xi}$ 按分布相等. 由于这一性质, 像几乎必然收敛与平均收敛一样, 只与概率分布有关, 则对于某个随机变量 $\widetilde{\eta}$, 由以概率 1 收敛与平均收敛意义下

$$\frac{1}{n}\sum_{k=1}^{n}\widetilde{\xi}_1(\widetilde{T}^{k-1}\widetilde{\omega}) \to \widetilde{\eta},$$

可得 $n^{-1}\sum\limits_{k=1}^{n}\xi_k(\omega)$ 也以概率 1 收敛与平均收敛意义下收敛于某随机变量 η: $\eta \overset{d}{=} \widetilde{\eta}$. 由定理 1 可见, 如果 $\widetilde{\mathbf{E}}|\widetilde{\xi}_1| < \infty$, 则 $\widetilde{\eta} = \widetilde{\mathbf{E}}(\widetilde{\xi}_1|\widetilde{\mathscr{S}})$, 其中 $\widetilde{\mathscr{S}}$ 是不变集合的全体 ($\widetilde{\mathbf{E}}$ 是对测度 $\widetilde{\mathbf{P}}$求平均). 现在描绘随机变量 η的结构.

定义 1 称集合 $A \in \mathscr{F}$ 相对于序列 ξ.为*不变的*, 如果存在集合 $B \in \mathscr{B}(\mathbb{R}^\infty)$, 使对于任何 $n \geqslant 1$, 有

$$A = \{\omega : (\xi_n, \xi_{n+1}, \cdots) \in B\}.$$

这样不变集合的全体构成 $\sigma-$ 代数, 记作 $\mathscr{S}_\xi$.

定义 2 平稳序列 ξ 称为*遍历的*, 如果任意不变集合的测度只有 0 或 1 两个可能值.

现在证明, 当 $n \to \infty$ 时, 随机变量 η 在以概率 1 收敛与平均收敛意义下, 是随机变量 $n^{-1}\sum\limits_{k=1}^{n}\xi_k(\omega)$ 的极限, 且可以取作 $\mathbf{E}(\xi_1|\mathscr{S})$. 为此首先注意到, 自然可以设

$$\eta(\omega) = \overline{\lim_n}\frac{1}{n}\sum_{k=1}^{n}\xi_n(\omega). \tag{7}$$

由上极限 $\overline{\lim}$ 的定义可见, 对于这样定义的随机变量 $\eta(\omega)$, 集合 $\{\omega : \eta(\omega) < y\} (y \in \mathbb{R})$ 是不变的, 从而 η 为 $\mathscr{S}_\xi-$ 可测. 其次, 设 $A \in \mathscr{S}_\xi$. 那么, 由于

$$\mathbf{E}\left|\frac{1}{n}\sum_{k=1}^{n-1}\xi_k - \eta\right| \to 0, \quad n \to \infty,$$

可见对于由 (7) 式定义的 η, 有

$$\frac{1}{n}\sum_{k=1}^{n}\int_A \xi_k \mathrm{d}\mathbf{P} \to \int_A \eta \mathrm{d}\mathbf{P}. \tag{8}$$

设 $B \in \mathscr{B}(\mathbb{R}^\infty)$ 满足条件: 对于任意 $k \geqslant 1$, 有 $A = \{\omega : (\xi_k, \xi_{k+1}, \cdots) \in B\}$. 那么, 由于 ξ 的平稳性, 可见

$$\int_A \xi_k \mathrm{d}\mathbf{P} = \int_{\{\omega:(\xi_k,\xi_{k+1},\cdots)\in B\}} \xi_k \mathrm{d}\mathbf{P} = \int_{\{\omega:(\xi_1,\xi_2,\cdots)\in B\}} \xi_1 \mathrm{d}\mathbf{P} = \int_A \xi_1 \mathrm{d}\mathbf{P}.$$

因此, 由 (8) 式可见, 对于任意 $A \in \mathscr{S}_\xi$, 有如下等式

$$\int_A \xi_1 \mathrm{d}\mathbf{P} = \int_A \eta \mathrm{d}\mathbf{P}.$$

该式表明 (见第二章 §7 的 (1) 式), ($\mathscr{S}_\xi$- 可测) 随机变量 $\eta = \mathbf{E}(\xi_1|\mathscr{S}_\xi)$, 并且如果序列 ξ 是遍历的, 则 $\mathbf{E}(\xi_1|\mathscr{S}_\xi) = \mathbf{E}\xi_1$.

于是, 证明了下面的定理.

定理 3 (遍历性定理) *设 $\xi = (\xi_1, \xi_2, \cdots)$ 是强平稳序列, $\mathbf{E}|\xi_1| < \infty$. 那么, (在以概率 1 收敛与平均收敛意义下)*

$$\lim_n \frac{1}{n} \sum_{k=1}^n \xi_k(\omega) = \mathbf{E}(\xi_1|\mathscr{S}_\xi).$$

并且, 如果同时 ξ 又是遍历序列, 则 (在以概率 1 收敛与平均收敛意义下)

$$\lim_n \frac{1}{n} \sum_{k=1}^n \xi_k(\omega) = \mathbf{E}\xi_1.$$

4. 练习题

1. 设 $\xi = (\xi_1, \xi_2, \cdots)$ 是高斯平稳序列, $\mathbf{E}\xi_n = 0$, 而相关函数为 $R(n) = \mathbf{E}\xi_{k+n}\xi_k$. 证明, $R(n) \to 0$ 的充分条件是, 对应于随机变量序列 ξ 的保测变换为混合 (因而是遍历的).

2. 对于任何独立同分布随机变量序列 $\xi = (\xi_1, \xi_2, \cdots)$, 相应的保测变换都是混合.

3. 证明平稳序列 ξ 为遍历的, 当且仅当对于任意 $B \in \mathscr{B}(\mathbb{R}^k), k = 1, 2, \cdots$, 以概率 1,

$$\frac{1}{n} \sum_{i=1}^n I_B(\xi_i, \cdots, \xi_{i+k-1}) \to \mathbf{P}\{(\xi_1, \cdots, \xi_k) \in B\}.$$

4. 设 $\mathbf{P}$ 和 $\overline{\mathbf{P}}$ 是 $(\Omega, \mathscr{F})$ 上的两个概率测度, 且关于 $\mathbf{P}$ 和 $\overline{\mathbf{P}}$ 的保测变换 T 是遍历的. 证明, 这时要么 $\mathbf{P} = \overline{\mathbf{P}}$, 要么 $\mathbf{P} \perp \overline{\mathbf{P}}$.

5. 设 T 是 $(\Omega, \mathscr{F}, \mathbf{P})$ 上的保测变换, 而 $\mathscr{A}$ 是 Ω 的子集的代数, 且 $\sigma(\mathscr{A}) = \mathscr{F}$; 设

$$I_A^{(n)} = \frac{1}{n} \sum_{k=0}^{n-1} I_A(T^k \omega).$$

证明变换 T 是遍历的, 当且仅当满足下列条件之一:

(a) 对于任意 $A \in \mathscr{A}, I_A^{(n)} \xrightarrow{\mathbf{P}} \mathbf{P}(A)$;

(b) 对于一切 $A, B \in \mathscr{A}$,

$$\lim \frac{1}{n}\sum_{k=0}^{n-1} \mathbf{P}(A \bigcap T^{-k}B) = \mathbf{P}(A)\mathbf{P}(B);$$

(c) 对于任意 $A \in \mathscr{F}, I_A^{(n)} \xrightarrow{\mathbf{P}} \mathbf{P}(A)$.

6. 设 T 是 $(\Omega, \mathscr{F}, \mathbf{P})$ 上的保测变换. 证明变换 T (关于测度 $\mathbf{P}$)是遍历的, 当且仅当在 $(\Omega, \mathscr{F})$ 上不存在概率测度 $\overline{\mathbf{P}} \neq \mathbf{P}, \overline{\mathbf{P}} \ll \mathbf{P}$, 而且 T 关于测度 $\overline{\mathbf{P}}$ 是保测变换.

7. (伯努利推移变换) 假设 S 是一有限集合 (例如, $S = \{1, 2, \cdots, N\}$), 而 $\Omega = S^\infty$ 是序列 $\omega = (\omega_0, \omega_1, \cdots)$ 的全体, 其中 $\omega_i \in S$. 设 $\xi_k(\omega) = \omega_k$, 并且定义推移变换: $T(\omega_0, \omega_1, \cdots) = (\omega_1, \omega_2, \cdots)$, 即对于 $\xi_0, \xi_1, \cdots$, 如果 $\xi_k(\omega) = \omega_k$, 则 $\xi_k(T\omega) = \omega_{k+1}$. 假设在集合 $S = \{1, 2, \cdots, N\}$ 的元素 i 上给定正数 p_i, 且 $p_1 + p_2 + \cdots + p_N = 1$ (即数组 $(p_1, p_2, \cdots, p_N)$ 是概率分布). 利用这一变换可以在 $(S^\infty, \mathscr{B}(S^\infty))$ 上定义测度 $\mathbf{P}$ (见第二章 §3)

$$\mathbf{P}\{\omega : (\omega_1, \cdots, \omega_k) = (u_1, \cdots, u_k)\} = p_{u_1} \cdots p_{u_k}.$$

换句话说, 按 $\xi_0(\omega), \xi_1(\omega), \cdots$ 独立的原则引进概率. 关于这样建立的测度 $\mathbf{P}$, 推移变换 T 习惯上称做伯努利推移或伯努利变换.

证明伯努利变换具有混合性质.

8. 设 T 是 $(\Omega, \mathscr{F}, \mathbf{P})$ 上的保测变换. 引进记号 $T^{-n}\mathscr{F} = \{T^{-n}A : A \in \mathscr{F}\}$, 并且称 $\sigma-$ 代数

$$\mathscr{F}_{-\infty} = \bigcap_{n=1}^{\infty} T^{-n}\mathscr{F}$$

是平凡的 ($\mathbf{P}-$ 平凡的), 如果中的每一个集合的测度为 0 或 1 (这样的变换称做柯尔莫戈洛夫变换). 证明柯尔莫戈洛夫变换具有遍历性, 并且具有混合性.

9. 设 $1 \leqslant p < \infty, T$ 是概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 上的保测变换, 而随机变量 $\xi(\omega) \in L^p(\Omega, \mathscr{F}, \mathbf{P})$.

证明空间 $L^p(\Omega, \mathscr{F}, \mathbf{P})$ 上的如下 (冯・诺伊曼 [J. von Neuman]) 遍历性定理:

$$\mathbf{E}\left|\frac{1}{n}\sum_{k=0}^{n-1} \xi(T^k\omega) - \eta(\omega)\right|^p \to 0, \quad n \to \infty.$$

10. 根据博雷尔定理 (第四章 §3 例 2): 区间 $[0, 1)$ 上数 ω 的二进制分解中 "1 的比率" 和 "0 的比率" (关于勒贝格测度) 几乎必然收敛于 1/2. 如果将这一结果视为由公式

$$T(\omega) = 2\omega (\text{mod } 1)$$

定义的区间 $[0,1)$ 到自身的变换 $T:[0,1)\to[0,1)$, 试利用遍历性定理 1, 证明上述博雷尔定理.

11. 同上题一样, 设 $\omega\in[0,1)$. 考虑由公式

$$T(\omega)=\begin{cases}0, & 若\ \omega=0,\\ \left\{\dfrac{1}{\omega}\right\}, & 若\ \omega\neq 0\end{cases}$$

定义的变换 $T:[0,1)\to[0,1)$, 其中 $\{x\}$ 表示数 x 的分数部分.

设 $P=P(\cdot)$ 是区间 $[0,1)$ 上的高斯测度, 由如下公式定义:

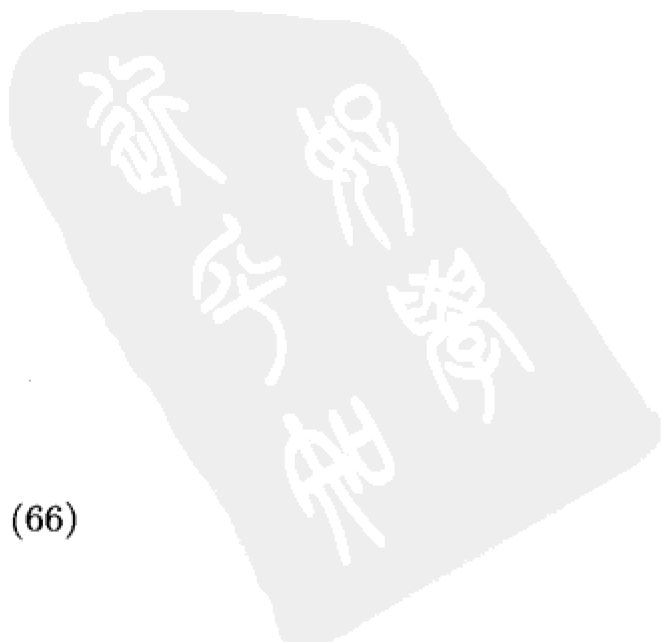
$$P(A)=\frac{1}{\ln 2}\int_A\frac{\mathrm{d}x}{1+x},\quad A\in\mathscr{B}([0,1)).$$

证明变换 T 保持高斯测度不变.

12. 举例说明, 对于具有无限测度的可测空间, 庞加莱 "常返定理" (§1 第 3 小节) 一般不成立.



第六章　弱 (广义) 平稳随机序列. L^2 理论





谱表示在弱平稳随机过程理论的中心位置，是研究弱平稳随机过程的基础. 任何平稳过程可看成, 具有随机振幅和相位的、各种频率的互不相关的调和振动的叠加.

《数学百科全书》(中译本) 第 4 卷, 第 914 页 [121]

---

§1. 协方差函数的谱表示

1. 平稳过程的概念与协方差函数 根据上一章给出的定义, 随机序列 $\xi = (\xi_1, \xi_2, \cdots)$ 称做强平稳序列①, 如果对于任意集合 $B \in \mathscr{B}(\mathbb{R}^\infty)$ 与任何 $n \geqslant 1$,

$$\mathbf{P}\{(\xi_1, \xi_2, \cdots) \in B\} = \mathbf{P}\{(\xi_{n+1}, \xi_{n+2}, \cdots) \in B\}. \tag{1}$$

特别, 由此可见, 若 $\mathbf{E}\xi_1^2 < \infty$, 则 $\mathbf{E}\xi_n$ 不依赖于 n:

$$\mathbf{E}\xi_n = \mathbf{E}\xi_1, \tag{2}$$

而且协方差 $\operatorname{cov}(\xi_{n+m}, \xi_n) = \mathbf{E}(\xi_{n+m} - \mathbf{E}\xi_{n+m})(\xi_n - \mathbf{E}\xi_n)$ 只依赖于 m:

$$\operatorname{cov}(\xi_{n+m}, \xi_n) = \operatorname{cov}(\xi_{1+m}, \xi_1). \tag{3}$$

这一章将研究 (具有有限二阶矩的) 所谓弱 (广义) 平稳随机序列, 这里将条件 (1) 换为条件 (2) 和 (3).

假设所考虑的随机变量 ξ_n, 对于 $n \in \mathbb{Z} = \{0, \pm 1, \cdots\}$ 有定义, 而且取复数为值. 这一假设非但不使理论复杂化, 反而使之更加雅致. 这时, 实值随机变量的有关结果, 自然容易由复值随机变量的相应结果, 作为其特殊情形得到.

设 $H^2 = H^2(\Omega, \mathscr{F}, \mathbf{P})$ 是 (复值) 随机变量 ξ 的值空间: $\xi = \alpha + i\beta, \alpha, \beta \in \mathbb{R}$, 且$\mathbf{E}|\xi|^2 < \infty$, 其中 $|\xi|^2 = \alpha^2 + \beta^2$. 如果 $\xi, \eta \in H^2$, 则设

$$(\xi, \eta) = \mathbf{E}\xi\overline{\eta}, \tag{4}$$

其中 $\overline{\eta} = \alpha - i\beta$ 是 $\eta = \alpha + i\beta$ 的共轭随机变量. 记

$$\|\xi\| = (\xi, \xi)^{1/2}. \tag{5}$$

①强平稳序列 (过程), 亦称严格意义下的平稳序列 (过程), 或严平稳序列 (过程). 弱平稳序列 (过程), 亦称广泛意义下的平稳序列 (过程), 或宽平稳序列 (过程). —— 译者

确切一点说, H^2 是等价随机变量类的空间 (见第二章 §10 和 §11). 像实值随机变量一样, 引进了内积 (ξ,η) 和范数 $\|\xi\|$ 的空间 H^2 是*完备的*. 按泛函分析的术语, 空间 H^2 称做所考虑概率空间 $(\Omega,\mathscr{F},\mathbf{P})$ 上的随机变量的*酉* (或*复*) *希尔伯特* (D. Hilbert) *空间*.

如果 $\xi,\eta\in H^2$, 则称

$$\operatorname{cov}(\xi,\eta)=\mathbf{E}(\xi-\mathbf{E}\xi)\overline{(\eta-\mathbf{E}\eta)} \tag{6}$$

为 ξ 和 η 的*协方差*.

由 (4) 和 (6) 式可见, 如果 $\mathbf{E}\xi=\mathbf{E}\eta=0$, 则

$$\operatorname{cov}(\xi,\eta)=(\xi,\eta). \tag{7}$$

定义 复值随机变量序列 $\xi=(\xi_n)_{n\in\mathbb{Z}},\mathbf{E}|\xi_n|^2<\infty,n\in\mathbb{Z}$, 称做*弱广义平稳的*, 如果对于所有 $n\in\mathbb{Z}$, 有

$$\begin{gathered}\mathbf{E}\xi_n=\mathbf{E}\xi_0,\\ \operatorname{cov}(\xi_{n+k},\xi_n)=\operatorname{cov}(\xi_n,\xi_0),\quad k\in\mathbb{Z}.\end{gathered} \tag{8}$$

为叙述简便, 以后总是假设 $\mathbf{E}\xi_0=0$. 这一假设并不影响一般性, 然而 (由于 (7) 式) 却可以使协方差与数积有相同的形式, 因此使希尔伯特空间理论的方法和结果用起来更加简单.

记

$$R(n)=\operatorname{cov}(\xi_n,\xi_0),\quad n\in\mathbb{Z}, \tag{9}$$

和 (在假设 $R(0)=\mathbf{E}|\xi_0|^2\neq 0$ 的条件下)

$$\rho(n)=\frac{R(n)}{R(0)},\quad n\in\mathbb{Z}. \tag{10}$$

我们把函数 $R(n)$ 称做 (弱平稳序列的) *协方差函数*, 而 $\rho(n)$ 称做*相关函数*.

直接由定义 (9) 可见, 协方差函数 $R(n)$ 是*非负定函数*, 即对于任意复数 $a_1,\cdots,a_m$, 与任何 $t_1,\cdots,t_m\in\mathbb{Z},m\geqslant 1$, 有

$$\sum_{i,j=1}^{m}a_i\overline{a}_jR(t_i-t_j)\geqslant 0. \tag{11}$$

同样, 由此 (或由 (9) 式) 不难得到 (练习题 1) 如下协方差函数的性质:

$$\begin{gathered}R(0)\geqslant 0,\quad R(-n)=\overline{R(n)},\quad |R(n)|\leqslant R(0),\\ |R(n)-R(m)|^2\leqslant 2R(0)[R(0)-\operatorname{Re}R(n-m)].\end{gathered} \tag{12}$$

2. 平稳序列的例 下面是一些平稳序列 $\xi = (\xi_n)_{n\in\mathbb{Z}}$ 的例. (为简便计, 下面常省略弱平稳过程的 "弱" 字, 以及$n \in \mathbb{Z}$).

例 1 设 $\xi_n = \xi_0 g(n)$, 其中 $\mathbf{E}\xi_0 = 0, \mathbf{E}|\xi_0|^2 = 1$, 而 $g = g(n)$ 是某一函数. 随机变量序列 $\xi = (\xi_n)$ 是平稳的, 当且仅当函数 $g(k+n)\overline{g(k)}$ 仅依赖于 n. 由此不难看到, 存在这样的 λ, 使

$$g(n) = g(0)\mathrm{e}^{-i\lambda n}.$$

因此, 随机变量序列

$$\xi_n = \xi_0 g(0)\mathrm{e}^{i\lambda n}$$

是平稳的, 且其协方差函数为

$$R(n) = |g(0)|^2\mathrm{e}^{i\lambda n}.$$

特别, "在时间上随机的常数" $\xi_n \equiv \xi_0$ 形成平稳序列.

注 我们关于例 1 指出, 由于 $\mathrm{e}^{i\lambda n} = \mathrm{e}^{in(\lambda+2\pi k)}, k = \pm1, \pm2, \cdots$, 则 (圆) 频率 λ 的值仅精确到任意被加项的 2π 倍. 按照习惯, 以后总假定 $\lambda \in [-\pi, \pi)$.

例 2 *殆周期性序列* 设

$$\xi_n = \sum_{k=1}^{N} z_k \mathrm{e}^{i\lambda_k n}, \quad n \in \mathbb{Z}, \tag{13}$$

其中 $z_1, \cdots, z_N$ 是正交随机变量: $\mathbf{E}z_i\bar{z}_j = 0, i \neq j$, 而且均值 $\mathbf{E}z_k = 0; \mathbf{E}|z_k|^2 = \sigma_k^2 > 0, -\pi \leqslant \lambda_k < \pi, k = 1, \cdots, N; \lambda_i \neq \lambda_j (i \neq j)$.

因此, 随机变量序列 $\xi = (\xi_n)$ 是平稳的, 且其协方差函数为

$$R(n) = \sum_{k=1}^{N} \sigma_k^2 \mathrm{e}^{i\lambda_k n}. \tag{14}$$

为推广 (13) 式, 现在设

$$\xi_n = \sum_{k=-\infty}^{\infty} z_k \mathrm{e}^{i\lambda_k n}, \tag{15}$$

其中 $z_k, k \in \mathbb{Z}$, 具有与 (13) 式中同样的性质. 如果假设 $\sum\limits_{k=-\infty}^{\infty} \sigma_k^2 < \infty$, 则 (15) 式右侧的级数均方收敛, 并且

$$R(n) = \sum_{k=-\infty}^{\infty} \sigma_k^2 \mathrm{e}^{i\lambda_k n}. \tag{16}$$

引进函数

$$F(\lambda) = \sum_{\{k:\lambda_k \leqslant \lambda\}} \sigma_k^2. \tag{17}$$

那么, 协方差函数 (16) 可以写成勒贝格－斯蒂尔切斯积分的形式

$$R(n)=\int_{-\pi}^{\pi}\mathrm{e}^{i\lambda n}\mathrm{d}F(\lambda)\left(=\int_{[-\pi,\pi)}\mathrm{e}^{i\lambda n}\mathrm{d}F(\lambda)\right). \tag{18}$$

平稳序列 (15) 是由 "谐波" $\mathrm{e}^{i\lambda_k n}$ 之和构成, 其 "频率" 为 λ_k, 而其随机 "振幅" z_k 的 "强度" 为 $\sigma_k^2=\mathbf{E}|z_k|^2$. 这样, 函数值 $F(\lambda)$ 关于序列 ξ 的 "谱" 的结构, 即关于构成表现 (15) 式之各种不同频率的强度值, 提供全面的信息. 根据 (18) 式, 函数 $F(\lambda)$ 的值亦完全决定协方差函数 $R(n)$ 的构造.

精确到常数因子, (非退化) 函数 $F(\lambda)$ 显然是分布函数, 并且在该例中函数 $F(\lambda)$ 是阶梯函数. 特别值得注意的是, 任意弱平稳序列的协方差函数可以表示为 (18) 式 (见第 3 小节的定理), 其中 $F(\lambda)$ (精确到规范因子) 是分布函数, 其承载子, 集中在区间 $[-\pi,\pi)$ 上, 即

$$F(\lambda)=\begin{cases}F(\pi-), & \text{若 } \lambda\geqslant\pi,\\ 0, & \text{若 } \lambda<-\pi.\end{cases}$$

由 (15) 和 (16) 两式形成的、协方差函数 $R(n)$ 的积分表示结果, 使我们产生一种想法: *任意*平稳序列都可以用 "积分" 表示. 实际上也正是这样, 在 §3 利用按正交测度的随机积分将要证明这一点.

例 3 *白噪声* 设 $\varepsilon=(\varepsilon_n)$ 是正交规范随机变量序列, $\mathbf{E}\varepsilon_n=0, \mathbf{E}\varepsilon_i\varepsilon_j=\delta_{ij}$, 其中 δ_{ij} 是克罗内克符号. 显然 $\varepsilon=(\varepsilon_n)$ 是平稳序列, 其协方差函数为

$$R(n)=\begin{cases}1, & \text{若 } n=0,\\ 0, & \text{若 } n\neq 0.\end{cases}$$

注意, 函数 $R(n)$ 可以表示为

$$R(n)=\int_{-\pi}^{\pi}\mathrm{e}^{i\lambda n}\mathrm{d}F(\lambda), \tag{19}$$

其中

$$F(\lambda)=\int_{-\pi}^{\lambda}f(v)\mathrm{d}v,\quad f(\lambda)=\frac{1}{2\pi},\quad -\pi\leqslant\lambda<\pi. \tag{20}$$

比较 "谱" 函数 (17) 和 (20) 式可见, 如果例 2 中的 "谱" 是*离散的*, 则在例 3 中它是*绝对连续的并有常数的谱密度* $f(\lambda)\equiv 1/2\pi$. 实际上可以说, 序列 $\varepsilon=(\varepsilon_n)$ "由具有同一强度的振动构成". 正是因为这个原因, 与由同一强度的不同色调形成的物理 "白噪声" 类似, 才把 $\varepsilon=(\varepsilon_n)$ 称做 "白噪声".

例 4 *移动平均序列* 由例 3 引进的白噪声 $\varepsilon=(\varepsilon_n)$ 出发, 组成新的序列

$$\xi_n=\sum_{k=-\infty}^{\infty}a_k\varepsilon_{n-k}, \tag{21}$$

其中 a_k 是复数, 满足 $\sum\limits_{k=-\infty}^{\infty}|a_k|^2<\infty$.

由 (21) 式, 可见

$$\operatorname{cov}(\xi_{n+m},\xi_m)=\operatorname{cov}(\xi_n,\xi_0)=\sum_{k=-\infty}^{\infty}a_{n+k}\overline{a}_k,$$

因此 $\xi=(\xi_k)$ 是平稳序列, 习惯上称做由序列 $\varepsilon=(\varepsilon_k)$ 形成的 (**双侧**) **移动平均序列**.

特别, 当所有带负下标的 a_k 都等于 0 时, 即

$$\xi_n=\sum_{k=0}^{\infty}a_k\varepsilon_{n-k},$$

$\xi=(\xi_n)$ 称做**单侧移动平均序列**. 与此同时, 若对于一切 $k>p, a_k=0$, 即如果

$$\xi_n=a_0\varepsilon_n+a_1\varepsilon_{n-1}+\cdots+a_p\varepsilon_{n-p}, \tag{22}$$

则 $\xi=(\xi_n)$ 称做 p **阶移动平均序列**.

可以证明 (练习题 3), 对于序列 (22), 其协方差函数 $R(n)$ 为:

$$R(n)=\int_{-\pi}^{\pi}\mathrm{e}^{i\lambda n}f(\lambda)\mathrm{d}\lambda,$$

其中谱密度 $f(\lambda)$ 为

$$f(\lambda)=\frac{1}{2\pi}|P(\mathrm{e}^{-i\lambda})|^2, \tag{23}$$

且

$$P(z)=a_0+a_1z+\cdots+a_pz^p.$$

例 5 **自回归概型** 设 $\varepsilon=(\varepsilon_n)$ 是白噪声. 称随机变量序列 $\xi=(\xi_n)$ 属于 q 阶**自回归概型**, 如果对于 $n\in\mathbb{Z}$, 有

$$\xi_n+b_1\xi_{n-1}+\cdots+b_q\xi_{n-q}=\varepsilon_n. \tag{24}$$

问当系数 $b_1,\cdots,b_q$ 满足何条件时, 方程 (24) 有平稳解? 为回答所提问题, 首先考虑 $q=1$ 的情形:

$$\xi_n=\alpha\xi_{n-1}+\varepsilon_n, \tag{25}$$

其中 $\alpha=-b_1$. 如果 $|\alpha|<1$, 则不难验证, 平稳序列 $\widetilde{\xi}=(\widetilde{\xi}_n)$:

$$\widetilde{\xi}_n=\sum_{j=0}^{\infty}\alpha^j\varepsilon_{n-j} \tag{26}$$

是方程 (25) 的解. (式 (26) 右侧的级数在均方意义下收敛.) 现在证明, 在具有有限二阶矩的平稳随机变量序列类中, $\widetilde{\xi}_n$ 是唯一解. 事实上, 由 (25) 式依次叠代, 得

$$\xi_n=\alpha\xi_{n-1}+\varepsilon_n=\alpha[\alpha\xi_{n-2}+\varepsilon_{n-1}]+\varepsilon_n=\cdots=\alpha^k\xi_{n-k}+\sum_{j=0}^{k-1}\alpha^j\varepsilon_{n-j}.$$

当 $k \to \infty$ 时, 由此得

$$\mathbf{E}\left[\xi_n - \sum_{j=0}^{k-1} \alpha^j \varepsilon_{n-j}\right]^2 = \mathbf{E}[\alpha^k \xi_{n-k}]^2 = \alpha^{2k}\mathbf{E}\xi_{n-k}^2 = \alpha^{2k}\mathbf{E}\xi_0^2 \to 0.$$

从而, 当 $|\alpha| < 1$ 时方程 (25) 的平稳解存在, 且表示为 (26) 式的单侧移动平均的形式.

对于任意 $q > 1$ 的情形, 有类似的结果: 假如多项式

$$Q(z) = 1 + b_1 z + \cdots + b_q z^q \tag{27}$$

的一切 0 点都位于单位圆之外, 则自回归方程有解, 而且有可表示为单侧移动平均的形式 (24) 的唯一平稳解 (练习题 2). 这时, 协方差函数 $R(n)$ 可以表示为 (练习题 3):

$$R(n) = \int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n} \mathrm{d}F(\lambda), \quad F(\lambda) = \int_{-\pi}^{\lambda} f(v)\mathrm{d}v, \tag{28}$$

其中

$$f(\lambda) = \frac{1}{2\pi} \times \frac{1}{|Q(\mathrm{e}^{-i\lambda})|^2}. \tag{29}$$

特别, 对于 $q = 1$ 的情形, 由 (25) 式容易得到: $\mathbf{E}\xi_0 = 0$,

$$\mathbf{E}|\xi_0|^2 = \frac{1}{1-|\alpha|^2}, \quad R(n) = \frac{\alpha^n}{1-|\alpha|^2}, \quad n \geqslant 0$$

(对于 $n < 0, R(n) = \overline{R(-n)}$). 这时

$$f(\lambda) = \frac{1}{2\pi} \times \frac{1}{|1-\alpha\mathrm{e}^{-i\lambda}|^2}.$$

例 6　这个例子表明, 在水文学中建立概率模型时产生自回归概型. 考虑某水域 (例如, 里海), 并设法建立描绘, 由于径流和水表面的蒸发产生的振动, 引起的水平面对平均水平之偏差的概率模型.

假如以年做计量单位, 以 H_n 表示第 n 年水域的 “水平”, 则可得如下平衡方程:

$$H_{n+1} = H_n - KS(H_n) + \Sigma_{\cdot n+1}, \tag{30}$$

其中以 $\Sigma_{\cdot n+1}$ 表示第 $(n+1)$ 年的径流量, 以 $S(H)$ 表示在水平为 H 时水域的表面积, 而 K 是蒸发系数.

以 $\xi_n = H_n - \overline{H}$ 表示对平均水平 $\overline{H}$ 的偏差 ($\overline{H}$ 是根据多年的观测结果估计得来的), 假设 $S(H) = S(\overline{H}) + c(H - \overline{H})$. 由平衡方程 (30), 可见随机变量满足方程:

$$\xi_{n+1} = \alpha\xi_n + \varepsilon_{n+1}, \tag{31}$$

其中 $\alpha = 1 - cK, \varepsilon_n = \Sigma_n - KS(\overline{H})$. 自然应认为随机变量 ε_n 的均值为 0, 并且作为初始逼近, 认为 $\{\varepsilon_n\}$是不相关和同分布的. 那么, 如同例 5 中曾证明的那样, (当 $|\alpha| < 1$ 时) 方程 (31) 有唯一平稳解, 应该将其视为所观测池中 (历年) 形成的水平面振动的状态.

作为可以由 (理论) 模型 (31) 得到的实际推论, 我们指出根据今年和往年的观测结果, 预测来年水面的水平的可能性. 结果恰好表明 (见下面 §6 的例 2), 由 $\cdots, \xi_{n-1}, \xi_n$ 的值对量 ξ_{n+1} 的最优均方线性估计, 就是 $\alpha\xi_n$.

例 7 自回归和移动平均的混合模型 如果在方程 (24) 的右侧将 ε_n 换成 $a_0\varepsilon_n + a_1\varepsilon_{n-1} + \cdots + a_p\varepsilon_{n-p}$, 则得所谓 (p,q) 阶自回归和移动平均的混合模型:

$$\xi_n + b_1\xi_{n-1} + \cdots + b_q\xi_{n-q} = a_0\varepsilon_n + a_1\varepsilon_{n-1} + \cdots + a_p\varepsilon_{n-p}. \tag{32}$$

在与例 5 相同的假设条件下, 关于多项式 $Q(z)$ 的 0 点, 下面将要证明 (§3 中定理 3 的系 2), 方程 (32) 有平稳解 $\xi = (\xi_n)$, 其协方差函数为:

$$R(n) = \int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n}\mathrm{d}F(\lambda), \quad F(\lambda) = \int_{-\pi}^{\lambda} f(v)\mathrm{d}v,$$

其中

$$f(\lambda) = \frac{1}{2\pi} \times \left|\frac{P(\mathrm{e}^{-i\lambda})}{Q(\mathrm{e}^{-i\lambda})}\right|^2.$$

3. 赫尔格洛茨定理

定理 (赫尔格洛茨 [G. Herglotz]) 设 $R(n)$ 是均值为 0 的 (弱) 平稳随机序列的协方差函数, 则在 $([-\pi,\pi), \mathscr{B}([-\pi,\pi))$ 上存在一有限测度 $F = F(B), B \in \mathscr{B}([-\pi,\pi))$, 使对于任何 $n \in \mathbb{Z}$, 有

$$R(n) = \int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n}F(\mathrm{d}\lambda), \tag{33}$$

其中的积分 $\int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n}F(\mathrm{d}\lambda)$ 是集合 $[-\pi,\pi)$ 上的勒贝格 – 斯蒂尔切斯积分.

证明 对于 $N \geqslant 1, \lambda \in [-\pi,\pi]$, 设

$$f_N(\lambda) = \frac{1}{2\pi N}\sum_{k=1}^{N}\sum_{l=1}^{N} R(k-l)\mathrm{e}^{-ik\lambda}\mathrm{e}^{-il\lambda}. \tag{34}$$

由于 $R(n)$ 的非负定性, 可见 $f_N(\lambda)$ 是非负函数. 由于对于 $k - l = m$, 数偶 (k,l) 的个数等于 $N - |m|$, 则

$$f_N(\lambda) = \frac{1}{2\pi}\sum_{|m|<N}\left(1 - \frac{|m|}{N}\right)R(m)\mathrm{e}^{-im\lambda}. \tag{35}$$

设

$$F_N(B) = \int_B f_N(\lambda)\mathrm{d}\lambda, \quad B \in \mathscr{B}([-\pi,\pi)),$$

则

$$\int_{-\pi}^{\pi} e^{i\lambda n} F_N(d\lambda) = \int_{-\pi}^{\pi} e^{i\lambda n} f_N(\lambda) d\lambda = \begin{cases} \left(1 - \dfrac{|n|}{N}\right) R(n), & 若\ |n| < N, \\ 0, & 若\ |n| \geqslant N. \end{cases} \tag{36}$$

测度 $F_N, N \geqslant 1$, 集中在区间 $[-\pi, \pi]$上, 且对于任何 $N \geqslant 1, F_N[-\pi, \pi] = R(0) < \infty$. 从而, 测度 $F_N, N \geqslant 1$, 的族完备. 可见根据*普罗霍罗夫* (Ю. В. Прохоров) 定理 (第三章 §2 定理 1), 存在数列 $\{N_k\} \subseteq \{N\}$ 和测度 F, 使 $F_{N_k} \xrightarrow{w} F$. (密度、相对列紧性、弱收敛等概念, 以及普罗霍罗夫定理, 显然可以由概率测度移植到任何有限测度).

那么, 由 (36) 式可见

$$\int_{-\pi}^{\pi} e^{i\lambda n} F(d\lambda) = \lim_{N_k \to \infty} \int_{-\pi}^{\pi} e^{i\lambda n} F_{N_k}(d\lambda) = R(n).$$

所建立的测度 F 集中在区间 $[-\pi, \pi]$ 上. 在不改变积分 $\displaystyle\int_{-\pi}^{\pi} e^{i\lambda n} F(d\lambda)$ 的情况下, 把集中在点 π 上的 "质量" $F(\{\pi\})$ 移到点 $-\pi$上, 可以*重新定义*测度 F. 这样得到的新测度 (仍记为 F) 就已经集中在区间 $[-\pi, \pi)$上. (关于恰好是在区间 $[-\pi, \pi)$ 上, 而不是在区间 $[-\pi, \pi]$ 上. 选择 λ 值的合理性, 见第 2 小节的例 1.) □

注 1 表达式 (33) 中的测度 $F = F(B)$, 称做*协方差函数*为 $R(n)$ 的平稳序列的*谱测度*, 而 $F(\lambda) = F([-\pi, \lambda])$ 称做*谱函数*.

在上面的例 2 中, 谱测度是离散的 (集中在点 $\lambda_k, k = 0, \pm 1, \cdots$). 在例 3 ~ 6 中谱测度是绝对连续的.

注 2 谱测度 F 由协方差函数唯一决定. 事实上, 设 F_1 和 F_2 是两个谱测度, 且

$$\int_{-\pi}^{\pi} e^{i\lambda n} F_1(d\lambda) = \int_{-\pi}^{\pi} e^{i\lambda n} F_2(d\lambda), \quad n \in \mathbb{Z}.$$

由于任何有界连续函数 $g(\lambda)$, 可以在区间 $[-\pi, \pi)$ 上由三角多项式逼近, 则

$$\int_{-\pi}^{\pi} g(\lambda) F_1(d\lambda) = \int_{-\pi}^{\pi} g(\lambda) F_2(d\lambda),$$

由此可见 (对照第二章 §12 定理 2 的证明), 对于任何 $B \in \mathscr{B}([-\pi, \pi)), F_1(B) = F_2(B)$.

注 3 如果 $\xi = (\xi_n)$ 是由*实随机变量* ξ_n 构成的平稳序列, 则 $R(n) = R(-n)$, 因而

$$R(n) = \frac{R(n) + R(-n)}{2} = \int_{-\pi}^{\pi} \cos \lambda n F(d\lambda).$$

4. 练习题

1. 由 (11) 式导出 (12) 式.

2. 证明, 如果 (27) 式的多项式 $Q(z)$ 的一切 0 点都位于单位圆之外, 则自回归方程 (24) 有且唯一表示为单侧移动平均的平稳解.

3. 证明序列 (22) 和 (24) 的谱函数有密度, 且分别表示为公式 (23) 和 (29).

4. 证明, 如果 $\sum_{n=-\infty}^{\infty}|R(n)|^2<\infty$, 则谱函数 $F(\lambda)$ 有密度 $f(\lambda)$, 且由如下公式:

$$f(\lambda)=\frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\mathrm{e}^{-in\lambda}R(n)$$

表示, 其中级数在复空间 $L^2=L^2([-\pi,\pi),\mathscr{B}([-\pi,\pi),\lambda))$中收敛, 而 λ 是勒贝格测度.

§2. 正交随机测度和随机积分

1. 扩充积分概念的必要性 在 §1 中已经讨论了协方差函数的积分表示, 并举出了具有两两正交随机变量 $z_k, k\in\mathbb{Z}$ 的平稳序列

$$\xi_n=\sum_{k=-\infty}^{\infty}z_k\mathrm{e}^{i\lambda_k n}, \tag{1}$$

的例. 因此, 使我们产生一种想法: 是否可以将任意平稳序列表示为, 和式 (1) 之相应积分推广的形式.

假如设

$$Z(\lambda)=\sum_{\{k:z_k\leqslant\lambda\}}z_k, \tag{2}$$

则 (1) 式可以写为

$$\xi_n=\sum_{k=-\infty}^{\infty}\mathrm{e}^{i\lambda_k n}\Delta Z(\lambda_k), \tag{3}$$

其中 $\Delta Z(\lambda_k)\equiv Z(\lambda_k)-Z(\lambda_k-)=z_k$.

式 (3) 的右侧是 “勒贝格 — 斯蒂尔切斯积分”

$$\int_{-\pi}^{\pi}\mathrm{e}^{i\lambda n}\mathrm{d}Z(\lambda)$$

的积分和. 然而, 对于现在所考虑的情形, 函数 $Z(\lambda)$ 也是 (依赖于 ω 的) 随机变量. 在这种情形下, 为说清楚任意平稳序列的积分表示, 不得不也考虑这样的函数 $Z(\lambda)$: 对于每一个 ω, $Z(\lambda)$ 有无界变差. 因此对于每一个 ω, 简单地把上述积分理解为黎曼 — 斯蒂尔切斯积分是不可行的.

2. 随机测度 与勒贝格积分, 勒贝格 — 斯蒂尔切斯积分以及黎曼 — 斯蒂尔切斯积分的一般概念类似 (第二章 §6) 从定义*随机测度*开始讨论我们所感兴趣的情形.

设 $(\Omega, \mathscr{F}, \mathbf{P})$ 是概率空间, E 是某一集合, $\mathscr{E}_0$ 是集合 E 子集的代数, 而 $\mathscr{E} = \sigma(\mathscr{E}_0)$ 是 $\sigma-$ 代数.

定义 1 对于 $\omega \in \Omega$ 和 $\Delta \in \mathscr{E}_0$, 定义的复数值函数 $Z(\Delta) = Z(\omega, \Delta)$, 称做*有限 - 可加随机测度*, 如果

1) 对于任意 $\Delta \in \mathscr{E}_0, \mathbf{E}|Z(\Delta)|^2 < \infty$;

2) 对于 $\mathscr{E}_0$ 中任意两个不相交集合 Δ_1 和 Δ_2.

$$Z(\Delta_1 + \Delta_2) = Z(\Delta_1) + Z(\Delta_2) \quad (\mathbf{P} - \text{a.c.}). \tag{4}$$

定义 2 有限 — 可加随机测度 $Z(\Delta)$, 称做*初等随机测度*, 如果对于 $\mathscr{E}_0$ 中任意两两不相交集合 $\Delta_1, \Delta_2, \cdots$, 其中 $\Delta = \Delta_1 + \Delta_2 + \cdots \in \mathscr{E}_0$, 有

$$\mathbf{E}\left|Z(\Delta) - \sum_{k=1}^{n} Z(\Delta_k)\right|^2 \to 0, \quad n \to \infty. \tag{5}$$

注 1 在 $\mathscr{E}_0$ 的集合上的初等随机测度的定义中, 假设随机测度的值属于希尔伯特空间 $H^2 = H^2(\Omega, \mathscr{F}, \mathbf{P})$, 而该测度在 (5) 式的均方意义上是可数可加的. 尚存在随机测度的其他定义, 在这些定义中不要求存在二阶矩, 而且也不要求测度在 (5) 式的均方意义上可数 - 可加性, 例如要求在依概率收敛或以概率 1 收敛的意义上可数 — 可加.

注 2 与非随机测度的情形类似, 可以证明: 对于有限 — 可加随机测度, "在 (5) 式的均方意义上的可数可加性", 等价于 "在均方意义上在 '0' 的连续性":

$$\mathbf{E}|Z(\Delta_n)|^2 \to 0, \quad \Delta_n \downarrow \varnothing, \quad \Delta_n \in \mathscr{E}_0. \tag{6}$$

在初等随机测度类中, 按下面定义的正交测度特别重要.

定义 3 初等随机测度 $Z(\Delta), \Delta \in \mathscr{E}_0$, 称做*正交测度* (或*具有正交值的测度*), 如果对于 $\mathscr{E}_0$ 中任意两个不相交集合 Δ_1 和Δ_2, 有

$$\mathbf{E}Z(\Delta_1)\overline{Z(\Delta_2)} = 0, \tag{7}$$

或者等价地, 如果对于 $\mathscr{E}_0$ 中任意两个集合 Δ_1 和 Δ_2, 有

$$\mathbf{E}Z(\Delta_1)\overline{Z(\Delta_2)} = \mathbf{E}|Z(\Delta_1 \cap \Delta_2)|^2. \tag{8}$$

记

$$m(\Delta) = \mathbf{E}|Z(\Delta)|^2, \quad \Delta \in \mathscr{E}_0. \tag{9}$$

易见, 对于初等正交随机测度, 集函数 $m = m(\Delta), \Delta \in \mathscr{E}_0$, 是有限测度, 从而根据卡拉泰奥多里 (C. Carathéodory) 定理 (第二章 §3), $m(\Delta)$ 可以延拓到 $(E, \mathscr{E})$ 上. 这样,

得到的测度仍然记作 $m = m(\Delta)$, 并称之为 (初等正交随机测度 $Z = Z(\Delta), \Delta \in \mathscr{E}_0$, 的) **构造函数**.

现在自然产生如下问题: 既然定义在 $(E, \mathscr{E}_0)$ 上的集函数 $m = m(\Delta)$ 可以延拓到 $(E, \mathscr{E})$ 上, 其中 $\mathscr{E} = \sigma(\mathscr{E}_0)$, 那么是否可以将初等正交随机测度 $Z = Z(\Delta), \Delta \in \mathscr{E}_0$, 延拓到集合 $\Delta \in \mathscr{E}$ 上, 并且使 $m(\Delta) = \mathbf{E}|Z(\Delta)|^2, \Delta \in \mathscr{E}$.

对该问题的回答是肯定的, 这可以由下面的构造证明. 与此同时由这一构造还可以建立, 平稳序列的积分表现所需要的随机积分.

3. 随机积分 设 $Z = Z(\Delta), \Delta \in \mathscr{E}_0$ 是初等正交随机测度, 而 $m = m(\Delta), \Delta \in \mathscr{E}_0$, 是其构造函数. 对于每一个只有有限个不同 (复数) 值的函数

$$f(\lambda) = \sum f_k I_{\Delta_k}(\lambda), \quad \Delta_k \in \mathscr{E}_0, \tag{10}$$

定义随机变量

$$\mathscr{S}(f) = \sum f_k Z(\Delta_k).$$

设 $L^2 = L^2(E, \mathscr{E}, m)$ 是希尔伯特复数值函数的空间, 其数积为

$$\langle f, g \rangle = \int_E f(\lambda)\overline{g(\lambda)}m(\mathrm{d}\lambda),$$

且范数为 $\|f\| = \langle f, f \rangle^{1/2}$, 而 $H^2 = H^2(\Omega, \mathscr{F}, \mathbf{P})$ 是希尔伯特复随机变量的空间, 相应的数积为

$$(\xi, \eta) = \mathbf{E}\xi\overline{\eta},$$

范数为 $\|\xi\| = \langle \xi, \xi \rangle^{1/2}$.

那么, 显然对于任意两个形如 (10) 式的函数 f 和 g, 有

$$(\mathscr{S}(f), \mathscr{S}(g)) = \langle f, g \rangle,$$

且

$$\|\mathscr{S}(f)\|^2 = \|f\|^2 = \int_E |f(\lambda)|^2 m(\mathrm{d}\lambda).$$

现在设 $f \in L^2$, 而 $\{f_n\}$ 是形如 (10) 式的函数, 满足 $\|f - f_n\| \to 0, n \to \infty$ (见练习题 2). 那么当 $n, m \to \infty$ 时, 有

$$\|\mathscr{S}(f_n) - \mathscr{S}(f_m)\| = \|f_n - f_m\| \to 0.$$

从而, 序列 $\{\mathscr{S}(f_n)\}$ 在均方意义上是基本的, 而由第二章 §10 的定理 7, 可见存在随机变量 (记作 $\mathscr{S}(f)$), 满足 $\mathscr{S}(f) \in H^2$ 且当 $n \to \infty$ 时 $\|\mathscr{S}(f_n) - \mathscr{S}(f)\| \to 0$.

这样建立的随机变量 $\mathscr{S}(f)$ (精确到随机等价性) 唯一, 而且与逼近序列 $\{f_n\}$ 的选择无关. 自然把 $\mathscr{S}(f)$ 称做函数 $f \in L^2$ 为关于初等正交随机测度 Z 的**随机积分**, 并且为了直观 (与 $\mathscr{S}(f)$ 同时) 使用 "积分" 记号:

$$\int_E f(\lambda)Z(\mathrm{d}\lambda).$$

现在指出随机积分 $\mathscr{S}(f)$ 的下列性质, 这些性质可以直接由其构造得出. 设函数 $g, f, f_n \in L^2$, 那么

$$(\mathscr{S}(f), \mathscr{S}(g)) = \langle f, g \rangle, \tag{11}$$

$$\|\mathscr{S}(f)\| = \|f\|, \tag{12}$$

$$\mathscr{S}(af + bg) = a\mathscr{S}(f) + b\mathscr{S}(g) \quad (\mathbf{P} - \text{a.c.}), \tag{13}$$

其中 a 和 b 是常数; 若当 $n \to \infty$ 时, $\|f_n - f\| \to 0$, 则

$$\|\mathscr{S}(f_n) - \mathscr{S}(f)\| \to 0. \tag{14}$$

4. 随机测度的延拓　将上面定义的随机积分, 用于初等正交测度 $Z(\Delta), \Delta \in \mathscr{E}_0$, 到 $\mathscr{E} = \sigma(\mathscr{E}_0)$ 中集合的**延拓**.

由于假设测度 m 是有限的, 则对于任意 $\Delta \in \mathscr{E}_0$, 函数 $I_\Delta = I_\Delta(\lambda) \in L^2$. 记 $\widetilde{Z}(\Delta) = \mathscr{S}(I_\Delta)$. 显然, 对于 $\Delta \in \mathscr{E}_0$, 有 $\widetilde{Z}(\Delta) = Z(\Delta)$. 由 (13) 式可见, 如果 $\Delta_1 \bigcap \Delta_2 = \varnothing, \Delta_1, \Delta_2 \in \mathscr{E}$, 则

$$\widetilde{Z}(\Delta_1 + \Delta_2) = \widetilde{Z}(\Delta_1) + \widetilde{Z}(\Delta_2) \quad (\mathbf{P} - \text{a.c.}),$$

而由 (12) 式, 得

$$\mathbf{E}|\widetilde{Z}(\Delta)|^2 = m(\Delta), \quad \Delta \in \mathscr{E}.$$

现在证明, 集合 $\widetilde{Z}(\Delta), \Delta \in \mathscr{E}$, 的随机函数, 是在均方意义上可数可加的. 事实上, 设 $\Delta_k \in \mathscr{E}, \Delta = \sum\limits_{k=1}^{\infty} \Delta_k$, 则

$$\widetilde{Z}(\Delta) - \sum_{k=1}^{n} \widetilde{Z}(\Delta_k) = \mathscr{S}(g_n),$$

其中

$$g_n(\lambda) = I_\Delta(\lambda) - \sum_{k=1}^{n} I_{\Delta_k}(\lambda) = I_\Sigma(\lambda), \quad \Sigma = \sum_{k=n+1}^{\infty} \Delta_k.$$

但是

$$\mathbf{E}|\mathscr{S}(g_n)|^2 = \|g_n\|^2 = m\left(\sum_{k=n+1}^{\infty} \Delta_k\right) \downarrow 0, \quad n \to \infty,$$

即

$$\mathbf{E}\left|\widetilde{Z}(\Delta) - \sum_{k=1}^{n} \widetilde{Z}(\Delta_k)\right|^2 \to 0, \quad n \to \infty.$$

由 (11) 式可见, 对于 $\Delta_1 \bigcap \Delta_2 = \varnothing, \Delta_1, \Delta_2 \in \mathscr{E}$, 有

$$\mathbf{E}\widetilde{Z}(\Delta_1)\overline{\widetilde{Z}(\Delta_2)} = 0.$$

这样所建立定义在集合 $\Delta \in \mathscr{E}$ 上的随机函数 $\widetilde{Z}(\Delta)$, 是在均方意义上可数可加的, 并且在集合 $\Delta \in \mathscr{E}_0$ 上等于 $Z(\Delta)$. 我们称 $\widetilde{Z}(\Delta), \Delta \in \mathscr{E}$ (作为初等正交随机测度$Z(\Delta)$ 的延拓) 是以 $m(\Delta), \Delta \in \mathscr{E}$, 为构造函数的*正交随机测度*, 而上面定义的积分 $\int_E f(\lambda) Z(\mathrm{d}\lambda)$ 称做关于该测度的*随机积分*.

5. 正交增量随机过程 现在考虑对我们的目的最重要的情形, 即 $(E, \mathscr{E}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 的情形. 由第二章 §3 定理 1 知, 空间 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 上的任何有限测度 $m = m(\Delta)$, 都与某一 (广义) 分布函数 $G = G(x)$ 相互一一对应, 并且 $m(a, b] = G(b) - G(a)$.

结果表明, 对于正交测度有同样的情形. 下面引进正交增量随机过程的概念.

定义 4 概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 上 (复数值) 随机变量 $\{Z_\lambda\}, \lambda \in \mathbb{R}$, 的全体, 称做*正交增量随机过程*, 如果

1) $\mathbf{E}|Z_\lambda|^2 < \infty, \lambda \in \mathbb{R}$;

2) 对于每一个 $\lambda \in \mathbb{R}$,

$$\mathbf{E}|Z_\lambda - Z_{\lambda_n}|^2 \to 0, \quad \lambda_n \downarrow \lambda, \quad \lambda_n \in \mathbb{R};$$

3) 对于任意 $\lambda_1 < \lambda_2 < \lambda_3 < \lambda_4$,

$$\mathbf{E}(Z_{\lambda_4} - Z_{\lambda_3})\overline{(Z_{\lambda_2} - Z_{\lambda_1})} = 0.$$

条件 3) 是增量的*正交性条件*. 条件 1) 表示 $Z_\lambda \in H^2$. 最后, 条件 2) 具有技术的特点, 表示要求在每一点 $\lambda \in \mathbb{R}$ (在均方意义上) 上是*右连续的*.

设 $Z = Z(\lambda)$ 是以 $m = m(\Delta)$ 为构造函数的正交随机测度, 是 (广义) 分布函数为 $G = G(\lambda)$ 的有限测度. 记

$$Z_\lambda = Z(-\infty, \lambda].$$

那么, $\mathbf{E}|Z_\lambda|^2 = m(-\infty, \lambda] = G(\lambda) < \infty, \mathbf{E}|Z_\lambda - Z_{\lambda_n}|^2 = m(\lambda_n, \lambda] \downarrow 0, \lambda_n \downarrow \lambda$, 并且条件 (3) 显然也成立. 因此, 所建立的过程 $Z(\lambda)$ 是*正交增量*过程.

另一方面, 如果 $G = G(\lambda)$ 是广义分布函数, 且 $G(-\infty) = 0, G(+\infty) < \infty$; 而 $\{Z_\lambda\}$ 是正交增量过程, 且 $\mathbf{E}|Z_\lambda|^2 = G(\lambda)$, 则对于 $\Delta \in (a, b]$, 设

$$Z(\Delta) = Z_b - Z_a.$$

设 $\mathscr{E}_0$ 是形如 $\Delta = \sum_{k=1}^{n} (a_k, b_k]$ 的集合产生的代数; 而 $Z(\Delta) = \sum_{k=1}^{n} Z(a_k, b_k]$. 显然

$$\mathbf{E}|Z(\Delta)|^2 = m(\Delta),$$

其中 $m(\Delta) = \sum_{k=1}^{n} [G(b_k) - G(a_k)]$, 而对于不相交区间 $\Delta_1 = (a_1, b_1]$ 和 $\Delta_2 = (a_2, b_2]$, 有

$$\mathbf{E}Z(\Delta_1)\overline{Z(\Delta_2)} = 0.$$

因为函数 $G(\lambda),\lambda\in\mathbb{R}$, 右连续, 故由此可见 $Z=Z(\Delta),\Delta\in\mathscr{E}_0$, 是具有正交值的初等随机测度. 集函数 $m=m(\Delta),\Delta\in\mathscr{E}_0$, 唯一地延拓到 $\mathscr{E}=\mathscr{B}(\mathbb{R})$ 上的测度, 而由前面的构造可见, 那么 $Z=Z(\Delta),\Delta\in\mathscr{E}_0$, 也可以延拓到 $\Delta\in\mathscr{E}=\mathscr{B}(\mathbb{R})$. 这时, 有

$$\mathbf{E}|Z(\Delta)|^2=m(\Delta),\quad \Delta\in\mathscr{B}(\mathbb{R}).$$

设 $\{Z_\lambda\},\lambda\in\mathbb{R}$, 是正交增量过程, $\mathbf{E}|Z_\lambda|^2=G(\lambda),G(-\infty)=0,G(+\infty)<\infty$; 而 $Z=Z(\Delta),\Delta\in\mathscr{B}(\mathbb{R})$, 是构造函数为 $m=m(\Delta)$ 的正交随机测度, 那么, 由以上的论述可见, 在 $\{Z_\lambda\}$ 与 $m=m(\Delta)$ 之间存在一一对应关系, 使

$$Z_\lambda=Z(-\infty,\lambda],\quad G(\lambda)=m(-\infty,\lambda]$$

和

$$Z(a,b]=Z_b-Z_a,\quad m(a,b]=G(b)-G(a).$$

根据在勒贝格 — 斯蒂尔切斯和黎曼 — 斯蒂尔切斯积分理论 (第二章 §6 第 9 和第 11 小节) 中通用的记号, 对于某一正交增量随机过程 $\{Z_\lambda\}$, 把随机积分 $\int_R f(\lambda)\mathrm{d}Z_\lambda$ 理解为与该过程 $\{Z_\lambda\}$ 相对应的、正交随机测度的随机积分 $\int_R f(\lambda)Z(\mathrm{d}\lambda)$.

6. 练习题

1. 证明条件 (5) 和 (6) 等价.

2. 设函数 $f\in L^2$. 利用第二章的结果 (第二章, §4 的定理 1, §6 定理 3 的系以及 §3 的练习题 8), 证明存在形如 (10) 式的函数序列 $\{f_n\}$, 使 $\|f_n-f\|\to 0,n\to\infty$.

3. 设 $Z(\Delta)$ 是以 $m(\Delta)$ 为构造函数的正交随机测度, 证明下列性质:

$$\mathbf{E}|Z(\Delta_1)-Z(\Delta_2)|^2=m(\Delta_1\triangle\Delta_2),$$
$$Z(\Delta_1\setminus\Delta_2)=Z(\Delta_1)-Z(\Delta_1\textstyle\bigcap\Delta_2)\quad(\mathbf{P}-\text{a.c.}),$$
$$Z(\Delta_1\triangle\Delta_2)=Z(\Delta_1)+Z(\Delta_2)-2Z(\Delta_1\textstyle\bigcap\Delta_2)\quad(\mathbf{P}-\text{a.c.}).$$

§3. 弱 (广义) 平稳序列的谱表示

1. 平稳随机序列谱表示　设 $\xi=(\xi_n)$ 是平稳随机序列, $\mathbf{E}\xi_n=0,n\in\mathbb{Z}$, 则根据 §1 的定理存在 $([-\pi,\pi),\mathscr{B}([-\pi,\pi)))$ 上的有限测度, 使协方差函数 $R(n)=\operatorname{cov}(\xi_{k+n},\xi_k)$ 有如下表现:

$$R(n)=\int_{-\pi}^{\pi}\mathrm{e}^{i\lambda n}F(\mathrm{d}\lambda).\tag{1}$$

下面的结果给出了序列 $\xi=(\xi_n),n\in\mathbb{Z}$, 的相应谱表示.

定理 1　*存在正交随机测度 $Z=Z(\Delta),\Delta\in\mathscr{B}([-\pi,\pi))$, 使对于每一个 $n\in\mathbb{Z}$, 以概率 1, 有*

$$\xi_n=\int_{-\pi}^{\pi}\mathrm{e}^{i\lambda n}Z(\mathrm{d}\lambda)\left(=\int_{[-\pi,\pi)}\mathrm{e}^{i\lambda n}Z(\mathrm{d}\lambda)\right).\tag{2}$$

这时 $\mathbf{E}Z(\Delta)=0, \mathbf{E}|Z(\Delta)|^2=F(\Delta)$.

证明 利用希尔伯特空间的有关结果, 进行证明最为简单.

设 $L^2(F)=L^2(E,\mathscr{E},F)$ 是复值函数的希尔伯特空间, $E=[-\pi,\pi), \mathscr{E}=\mathscr{B}([-\pi,\pi))$, 其中定义了内积

$$\langle f,g\rangle=\int_{-\pi}^{\pi} f(\lambda)\overline{g(\lambda)}F(\mathrm{d}\lambda), \tag{3}$$

而 $L_0^2(F)(L_0^2(F)\subseteq L^2(F))$ 是由 $e_n=e(\lambda), n\in\mathbb{Z}$, 生成的线性流形, 其中 $e_n(\lambda)=\mathrm{e}^{i\lambda n}$.

易见, 由于 $E=[-\pi,\pi)$, 而测度 F 有限, 可见流形 $L_0^2(F)$ 的闭包与 $L^2(F)$ 相等 (练习题 1):

$$\overline{L_0^2(F)}=L^2(F).$$

其次, 设 $L_0^2(\xi)$ 是由随机变量 $\xi_n, n\in\mathbb{Z}$, 生成的线性流形, 而 $L^2(\xi)(=\overline{L_0^2(\xi)})$ 是 (关于测度 $\mathbf{P}$ 的) 闭包.

我们现在建立随机元 $L_0^2(F)$ 与 $L_0^2(\xi)$ 之间一一对应 "$\leftrightarrow$" 关系,设

$$e_n\leftrightarrow\xi_n,\quad n\in\mathbb{Z}, \tag{4}$$

并且对于任意随机元 (确切地说, 对于任意等价随机元类), 按线性关系补充广义为

$$\sum\alpha_n e_n\leftrightarrow\sum\alpha_n\xi_n \tag{5}$$

(这里假设只有有限个复数 α_n 不为 0).

注意, 对应关系 (5) 式在如下意义上是适定的: 对于测度 F 几乎所有 $\sum\alpha_n e_n=0$, 当且仅当, 对于测度 $\mathbf{P}$ 几乎必然 $\sum\alpha_n\xi_n=0$.

这样定义的对应关系 "$\leftrightarrow$", 称做*等距的*, 即保持内积不变的. 事实上, 由 (3) 式, 有

$$\begin{aligned}\langle e_n,e_m\rangle&=\int_{-\pi}^{\pi}e_n(\lambda)\overline{e_m(\lambda)}F(\mathrm{d}\lambda)=\int_{-\pi}^{\pi}\mathrm{e}^{i\lambda(n-m)}F(\mathrm{d}\lambda)\\&=R(n-m)=\mathbf{E}\xi_n\bar{\xi}_m=(\xi_n,\xi_m),\end{aligned}$$

而且类似地有

$$\left\langle\sum\alpha_n e_n,\sum\beta_n e_n\right\rangle=\left(\sum\alpha_n\xi_n,\sum\beta_n\xi_n\right). \tag{6}$$

现在设 $\eta\in L^2(\xi)$. 由于 $L^2(\xi)=\overline{L_0^2}(\xi)$, 则存在随机变量序列 (η_n), 使 $\eta_n\in L_0^2(\xi)$ 和 $\|\eta_n-\eta\|\to 0, n\to\infty$. 从而 (η_n) 是基本随机变量序列, 说明函数 (f_n) 也是基本随机变量序列, 其中 $f_n\in L_0^2(F)$ 且 $f_n\leftrightarrow\eta_n$. 由于空间 $L^2(F)$ 是完备的, 因而存在函数 $f\in L^2(F)$, 使 $\|f_n-f\|\to 0, n\to\infty$.

显然相反的结果也成立: 如果 $f\in L^2(F)$, 且 $\|f_n-f\|\to 0, f_n\in L_0^2(F)$, 则存在这样的元素 $\eta\in L^2(\xi)$, 使 $\eta_n\in L_0^2(\xi)$ 和 $\|\eta_n-\eta\|\to 0, n\to\infty$ 且 $\eta_n\leftrightarrow f_n$.

至此, (等距) 对应关系 "$\leftrightarrow$", 暂时仅在 $L_0^2(\xi)$ 和 $L_0^2(F)$ 元素之间得以证明. 根据连续性, 将此对应关系补充定义: 设 $\eta \leftrightarrow f$, 其中 f 和 η 是上面所考虑的元素. 不难验证, 所建立的对应关系是 (等价随机变量类与函数类之间的) 线性的和保持内积不变的一一对应关系.

考虑函数 $f(\lambda) = I_\Delta(\lambda)$, 其中 $\Delta \in \mathscr{B}([-\pi,\pi)), \lambda \in [-\pi,\pi)$, 并设 $Z(\Delta) \in L^2(\xi)$ 是满足 $I_\Delta(\lambda) \leftrightarrow Z(\Delta)$ 的元素. 显然 $\|I_\Delta(\lambda)\|^2 = F(\Delta)$, 因而 $\mathbf{E}|Z(\Delta)|^2 = F(\Delta)$. 由于 $\mathbf{E}\xi_n = 0, n \in \mathbb{Z}$, 故对于每一个元素 $L_0^2(\xi)$ (从而, 对于 $L^2(\xi)$), 其数学期望等于 0. 特别 $\mathbf{E}Z(\Delta) = 0$. 此外, 如果 $\Delta_1 \bigcap \Delta_2 = \varnothing$, 则 $\mathbf{E}Z(\Delta_1)\overline{Z(\Delta_2)} = 0$ 且

$$\mathbf{E}\left|Z(\Delta) - \sum_{k=1}^{n} Z(\Delta_k)\right|^2 \to 0, \quad n \to \infty,$$

其中 $\Delta = \Delta_1 + \Delta_2 + \cdots$.

这样, 元素的全体 $Z = Z(\Delta), \Delta \in \mathscr{B}([-\pi,\pi))$ 本身是正交随机测度, 因此 (由于 §2) 由该测度可以定义随机积分

$$\mathscr{S}(f) = \int_{-\pi}^{\pi} f(\lambda) Z(\mathrm{d}\lambda), \quad f \in L^2(F).$$

设 $f \in L^2(F)$ 和 $\eta \leftrightarrow f$. 以 $\Phi(f)$ 表示元素 η (确切地说, 在每一等价随机变量类与函数类中, 选一个元素作为代表). 现在证明, 以概率 1, 有

$$\mathscr{S}(f) = \Phi(f). \tag{7}$$

事实上, 如果

$$f(\lambda) = \sum \alpha_k I_{\Delta_k}(\lambda) \tag{8}$$

是函数 $I_{\Delta_k}(\lambda), \Delta_k = (a_k, b_k]$, 的有限线性组合, 则根据随机积分的定义, 显然 $\mathscr{S}(f) = \sum \alpha_k Z(\Delta_k)$ 等于 $\Phi(f)$. 因此对于函数 (8), (7) 式成立. 但是, 如果 $f \in L^2(F)$ 且 $\|f_n - f\| \to 0$, 其中 f_n 是形如 (8) 式的函数, 则 $\|\Phi(f_n) - \Phi(f)\| \to 0$, 而根据 §2 的 (14) 式 $\|\mathscr{S}(f_n) - \mathscr{S}(f)\| \to 0$. 从而, 以概率 1 有 $\Phi(f) = \mathscr{S}(f)$.

取函数 $f(\lambda) = \mathrm{e}^{i\lambda n}$, 则由 (4) 式可见 $\Phi(\mathrm{e}^{i\lambda n}) = \xi_n$; 另一方面, $\mathscr{S}(f) = \int_{-\pi}^{\pi} f(\lambda) Z(\mathrm{d}\lambda)$. 于是, 由 (7) 式可见, 以概率 1, 有

$$\xi_n = \int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n} Z(\mathrm{d}\lambda), \quad n \in \mathbb{Z}.$$

□

系 1　设 $\xi = (\xi_n)$ 是平稳随机序列, 由实值随机变量 $\xi_n, n \in \mathbb{Z}$ 构成. 那么, 谱表达式 (2) 中的随机测度 $Z = Z(\Delta)$ 具有如下性质: 对于任何 $\Delta \in \mathscr{B}([-\pi,\pi))$,

$$Z(\Delta) = \overline{Z(-\Delta)}, \tag{9}$$

其中集合 $-\Delta = \{\lambda : -\lambda \in \Delta\}$.

事实上, 设 $f(\lambda) = \sum \alpha_k e^{i\lambda k}$ 和 $\eta = \sum \alpha_k \xi_k$ (两个和式都是有限的). 那么, $f \leftrightarrow \eta$, 因此

$$\overline{\eta} = \sum \overline{\alpha}_k \xi_k \leftrightarrow \sum \overline{\alpha}_k e^{i\lambda k} = \overline{f(-\lambda)}. \tag{10}$$

由于 $I_\Delta(\lambda) \leftrightarrow Z(\Delta)$, 则由 (10) 式, 可见 $I_\Delta(-\lambda) \leftrightarrow \overline{Z(\Delta)}$ (或等价地 $I_{-\Delta}(\lambda) \leftrightarrow \overline{Z(\Delta)}$). 另一方面, 因为 $I_{-\Delta}(\lambda) \leftrightarrow Z(-\Delta)$, 所以 $\overline{Z(\Delta)} = Z(-\Delta)(\mathbf{P}-\text{a.c.})$.

系 2 设 $\xi = (\xi_n)$ 是平稳随机序列, 其中 ξ_n 是实值随机变量, 而 $Z(\Delta) = Z_1(\Delta) + iZ_2(\Delta)$. 那么, 对于任意 $\Delta_1, \Delta_2 \in \mathscr{B}([-\pi, \pi))$, 有

$$\mathbf{E}Z_1(\Delta_1)Z_2(\Delta_2) = 0, \tag{11}$$

并且, 如果 $\Delta_1 \bigcap \Delta_2 = \varnothing$ 和 $(-\Delta_1) \bigcap \Delta_2 = \varnothing$, 则

$$\mathbf{E}Z_1(\Delta_1)Z_1(\Delta_2) = 0, \quad \mathbf{E}Z_2(\Delta_1)Z_2(\Delta_2) = 0. \tag{12}$$

事实上, 由于 $Z(\Delta) = \overline{Z(-\Delta)}$, 则

$$Z_1(-\Delta) = Z_1(\Delta), \quad Z_2(-\Delta) = -Z_2(\Delta). \tag{13}$$

其次, 因为 $\mathbf{E}Z(\Delta_1)\overline{Z(\Delta_2)} = \mathbf{E}|Z(\Delta_1 \bigcap \Delta_2)|^2$,

$$\mathrm{Im}\mathbf{E}Z(\Delta_1)\overline{Z(\Delta_2)} = 0,$$

即

$$\mathbf{E}Z_1(\Delta_1)Z_2(\Delta_2) - \mathbf{E}Z_2(\Delta_1)Z_1(\Delta_2) = 0. \tag{14}$$

将 Δ_1 换成区间 $-\Delta_1$, 则得

$$\mathbf{E}Z_1(-\Delta_1)Z_2(\Delta_2) - \mathbf{E}Z_2(-\Delta_1)Z_1(\Delta_2) = 0,$$

故由于 (13) 式, 由此得

$$\mathbf{E}Z_1(\Delta_1)Z_2(\Delta_2) + \mathbf{E}Z_2(\Delta_1)Z_1(\Delta_2) = 0. \tag{15}$$

由 (14) 和 (15) 式得等式 (11).

如果 $\Delta_1 \bigcap \Delta_2 = \varnothing$ 和 $(-\Delta_1) \bigcap \Delta_2 = \varnothing$, 则 $\mathbf{E}Z(\Delta_1)\overline{Z(\Delta_2)} = 0$. 因此 $\mathrm{Re}\,\mathbf{E}Z(\Delta_1)\overline{Z(\Delta_2)} = 0$ 和 $\mathrm{Re}\,\mathbf{E}Z(-\Delta_1)\overline{Z(\Delta_2)} = 0$. 于是, 连同 (13) 式就证明了等式 (12).

系 3 设 $\xi = (\xi_n)$ 是高斯序列. 那么, 对于任意 $\Delta_1, \cdots, \Delta_k$, 向量 $(Z_1(\Delta_1), \cdots, Z_1(\Delta_k); Z_2(\Delta_1), \cdots, Z_2(\Delta_k))$ 服从高斯 (正态) 分布.

事实上, 由 (复数值) 高斯随机变量 η, 即由服从高斯分布的向量 $(\mathrm{Re}\,\eta, \mathrm{Im}\,\eta)$ 生成的线性流形 $L_0^2(\xi)$, 服从高斯分布. 那么, 根据第二章 §13 第 5 小节, 可见 $L_0^2(\xi)$ 的闭包 $\overline{L_0^2(\xi)}$ 也由高斯随机变量构成. 因此由系 2 知, 对于高斯序列 $\xi = (\xi_n)$,

实部 Z_1 和虚部 Z_2 在如下意义上独立: 任意随机变量组 $(Z_1(\Delta_1),\cdots,Z_1(\Delta_k))$ 和 $(Z_2(\Delta_1),\cdots,Z_2(\Delta_k))$ 相互独立. 由 (12) 知, 对于满足 $\Delta_i\bigcap\Delta_j=(-\Delta_i)\bigcap\Delta_j=\varnothing, i,j=1,\cdots,k, i\neq j$, 的集合 $\Delta_1,\cdots,\Delta_k$, 随机变量 $Z_i(\Delta_1),\cdots,Z_i(\Delta_k)(i=1,2)$ 全体独立.

系 4 *设 $\xi=(\xi_n)$ 是平稳实值随机变量序列, 则以概率 1*

$$\xi_n=\int_{-\pi}^{\pi}\cos\lambda n Z_1(\mathrm{d}\lambda)-\int_{-\pi}^{\pi}\sin\lambda n Z_2(\mathrm{d}\lambda). \tag{16}$$

注 如果 $\{Z_\lambda\}, \lambda\in[-\pi,\pi)$ 是对应于正交随机测度 $Z=Z(\Delta)$ 的正交增量随机过程, 则谱表示 (2) (根据 §2) 可以表示为如下形式:

$$\xi_n=\int_{-\pi}^{\pi}\mathrm{e}^{i\lambda n}\mathrm{d}Z_\lambda,\quad n\in\mathbb{Z}. \tag{17}$$

2. 由平稳序列经线性变换得到的随机变量的构造 设 $\xi=(\xi_n)$ 是具有谱分解 (2) 的平稳序列, 而 $\eta\in L^2(\xi)$. 下面的定理描绘这样随机变量 η 的构造.

定理 2 *如果 $\eta\in L^2(\xi)$, 则存在函数 $\varphi\in L^2(F)$, 使以概率 1, 有*

$$\eta=\int_{-\pi}^{\pi}\varphi(\lambda)Z(\mathrm{d}\lambda). \tag{18}$$

证明 如果

$$\eta_n=\sum_{|k|\leqslant n}\alpha_k\xi_k, \tag{19}$$

则由于 (2) 式, 有

$$\eta_n=\int_{-\pi}^{\pi}\left(\sum_{|k|\leqslant n}\alpha_k\mathrm{e}^{i\lambda n}\right)Z(\mathrm{d}\lambda), \tag{20}$$

即对于函数

$$\varphi(\lambda)=\sum_{|k|\leqslant n}\alpha_k\mathrm{e}^{i\lambda n}, \tag{21}$$

(18) 式成立. 在一般情形下, 如果 $\eta\in L^2(\xi)$, 则存在形如 (19) 式的随机变量 η_n, 使 $\|\eta-\eta_n\|\to 0, n\to\infty$. 那么, $\|\varphi_n-\varphi_m\|=\|\eta_n-\eta_m\|\to 0, n,m\to\infty$, 即序列 (φ_n) 在 $L^2(F)$ 中是基本序列. 因此, 存在函数 $\varphi\in L^2(F)$, 使 $\|\varphi-\varphi_n\|\to 0, n\to\infty$.

根据 §2 的 (14) 式 $\|\mathscr{S}(\varphi_n)-\mathscr{S}(\varphi)\|\to 0$, 而因为 $\eta_n=\mathscr{S}(\varphi_n)$, 所以以概率 1, 有 $\eta=\mathscr{S}(\varphi)$. □

注 设 $H_0(\xi)$ 和 $H_0(F)$ 是相应为变量 $\xi^0=(\xi_n)_{n\leqslant 0}$ 和函数 $e^0=(e_n)_{n\leqslant 0}$ 封闭线性流形. 那么, 如果 $\eta\in H_0(\xi)$, 则存在这样一个函数 $\varphi\in H_0(F)$, 使以概率 1, 有

$$\eta=\int_{-\pi}^{\pi}\varphi(\lambda)Z(\mathrm{d}\lambda).$$

3. 线性滤波器 公式 (18) 描绘经线性变换, 由 $\xi_n, n \in \mathbb{Z}$ 得到的随机变量, 即可以表示为形如 (19) 的随机变量及其均方极限的构造.

利用所谓 (线性) 滤波器表示的线性变换类, 是特殊然而重要的线性变换类. 假设在某系统 (滤波器) 入口于时刻 m 出现信号 x_m, 而这时系统对该信号的反映是: 在其出口于时刻 n 收到信号 $h(n-m)x_m$, 其中 $h(s), s \in \mathbb{Z}$, 是某一复数值函数, 称为 (滤波器的) 脉冲转移函数.

这样, 在系统输出的和信号 y_n 可以表示为:

$$y_n = \sum_{m=-\infty}^{\infty} h(n-m)x_m. \tag{22}$$

对于物理可实现的系统, 输出信号的值只由 "过去" 的值决定, 即由 $x_m (m \leqslant n)$ 值决定. 如果对于一切 $s < 0$ 有 $h(s) = 0$, 或如果

$$y_n = \sum_{m=-\infty}^{\infty} h(n-m)x_m = \sum_{m=0}^{\infty} h(m)x_{n-m}, \tag{23}$$

则自然称脉冲转移函数为 $h = h(s)$ 的滤波器是物理可实现的.

称为滤波器的频率特征的、脉冲转移函数 h 的傅里叶变换

$$\varphi(\lambda) = \sum_{m=-\infty}^{\infty} \mathrm{e}^{-i\lambda m} h(m), \tag{24}$$

是以 h 为脉冲转移函数的滤波器的重要谱特征.

现在考虑关于 (22) 和 (24) 式中级数的收敛条件. 这些条件到目前为止尚未得到说明。假设在滤波器的入口发送具有协方差函数为 $R(n)$ 和谱分解 (2) 的、平稳随机序列 $h(s), s \in \mathbb{Z}$. 那么, 如果

$$\sum_{k,l=-\infty}^{\infty} h(k)R(l-k)\overline{h(l)} < \infty, \tag{25}$$

则级数 $\sum\limits_{m=-\infty}^{\infty} h(n-m)\xi_m$ 均方收敛, 因而平稳序列 $\eta = (\eta_n)$ 有定义, 其中

$$\eta_n = \sum_{m=-\infty}^{\infty} h(n-m)\xi_m = \sum_{m=-\infty}^{\infty} h(m)\xi_{n-m}. \tag{26}$$

按谱分析的术语, 条件 (25) 显然等价于 $\varphi(\lambda) \in L^2(F)$, 即

$$\int_{-\pi}^{\pi} |\varphi(\lambda)|^2 F(\mathrm{d}\lambda) < \infty. \tag{27}$$

在 (25) 式或 (27) 式的条件下, 由 (26) 式和 (2) 式, 可得序列 η 的谱表现

$$\eta_n = \int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n} \varphi(\lambda) Z(\mathrm{d}\lambda) < \infty, \quad n \in \mathbb{Z}. \tag{28}$$

从而, 序列 η 的协方差函数 $R_\eta(n)$ 由如下公式确定:

$$R_\eta(n) = \int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n} |\varphi(\lambda)|^2 F(\mathrm{d}\lambda) < \infty. \tag{29}$$

特别, 如果在频率特征为 $\varphi = \varphi(\lambda)$ 的滤波器入口传送白噪声 $\varepsilon = (\varepsilon_n)$, 则在其出口将收到平稳移动均值序列:

$$\eta_n = \sum_{m=-\infty}^{\infty} h(m)\varepsilon_{n-m}, \tag{30}$$

其谱密度为

$$f_\eta(\lambda) = \frac{1}{2\pi}|\varphi(\lambda)|^2.$$

下面的定理在一定意义上说明, 任何有谱密度的平稳序列可以由移动均值得到.

定理 3 *设 $\eta = (\eta_n)$ 是谱密度为 $f_\eta(\lambda)$ 的平稳序列, 则存在这样的白噪声序列 $\varepsilon = (\varepsilon_n)$ 和这样的滤波器 (必要时扩充原概率空间), 使表达式 (30) 成立.*

证明 由给定的 (非负) 函数 $f_\eta(\lambda)$, 存在这样的函数 $\varphi(\lambda)$, 使 $f_\eta(\lambda) = (2\pi)^{-1}|\varphi(\lambda)|^2$. 由于

$$\int_{-\pi}^{\pi} f_\eta(\lambda)\mathrm{d}\lambda < \infty, \quad 故 \quad \varphi(\lambda) \in L^2(\mathrm{d}\mu),$$

这里 $\mathrm{d}\mu$ 是 $[-\pi, \pi)$ 上的勒贝格测度. 因此, 函数 $\varphi(\lambda)$ 可以表示为傅里叶级数 (24), 其中

$$h(m) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{im\lambda}\varphi(\lambda)\mathrm{d}\lambda,$$

并且收敛性应理解为

$$\int_{-\pi}^{\pi} \left|\varphi(\lambda) - \sum_{|m|\leqslant n} \mathrm{e}^{-im\lambda}h(m)\right|^2 \mathrm{d}\lambda \to 0, \quad n \to \infty.$$

设

$$\eta_n = \int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n} Z(\mathrm{d}\lambda) < \infty, \quad n \in \mathbb{Z}.$$

除了测度 $Z = Z(\Delta)$ 之外, 我们再引进不依赖于 $Z = Z(\Delta)$ 的、新的正交随机测度 $\widetilde{Z} = \widetilde{Z}(\Delta)$, 并满足 $\mathbf{E}|\widetilde{Z}(a, b]|^2 = (b-a)/(2\pi)$. (一般, 建立这样测度的可能性, 应以原概率空间充分 "丰富" 为前提.) 设

$$\overline{Z}(\Delta) = \int_\Delta \varphi^{\oplus}(\lambda)Z(\mathrm{d}\lambda) + \int_\Delta [1 - \varphi^{\oplus}(\lambda)\varphi(\lambda)]\widetilde{Z}(\mathrm{d}\lambda),$$

其中

$$a^{\oplus} = \begin{cases} a^{-1}, & 若\ a \neq 0, \\ 0, & 若\ a = 0. \end{cases}$$

随机测度 $\overline{Z} = \overline{Z}(\Delta)$ 是具有正交值的测度, 这时对于任何 $\Delta = (a, b]$,

$$\mathbf{E}|\overline{Z}(\Delta)|^2 = \frac{1}{2\pi}\int_{\Delta} |\varphi^{\oplus}(\lambda)|^2|\varphi(\lambda)|^2 \mathrm{d}\lambda + \frac{1}{2\pi}\int_{\Delta} |1 - \varphi^{\oplus}(\lambda)\varphi(\lambda)|^2 \mathrm{d}\lambda = \frac{|\Delta|}{2\pi},$$

其中 $|\Delta| = b - a$. 因此, 平稳序列 $\varepsilon = (\varepsilon_n), n \in \mathbb{Z}$, 是白噪声, 其中

$$\varepsilon_n = \int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n}\overline{Z}(\mathrm{d}\lambda).$$

现在, 注意到,

$$\int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n}\varphi(\lambda)\overline{Z}(\mathrm{d}\lambda) = \int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n} Z(\mathrm{d}\lambda) = \eta_n; \tag{31}$$

而另一方面, 根据 $\varphi(\lambda)$ 的定义和 §2 的性质 (14), 以概率 1, 有

$$\begin{aligned}\int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n}\varphi(\lambda)\overline{Z}(\mathrm{d}\lambda) &= \int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n}\left(\sum_{m=-\infty}^{\infty} \mathrm{e}^{-i\lambda m} h(m)\right)\overline{Z}(\mathrm{d}\lambda) \\ &= \sum_{m=-\infty}^{\infty} h(m)\int_{-\pi}^{\pi} \mathrm{e}^{i\lambda(n-m)}\overline{Z}(\mathrm{d}\lambda) = \sum_{m=-\infty}^{\infty} h(m)\varepsilon_{n-m},\end{aligned}$$

于是, 注意到 (31) 式就证明了表达式 (30). □

注 如果 (按勒贝格测度几乎处处) $f_\eta(\lambda) > 0$, 则引进辅助测度 $\widetilde{Z} = \widetilde{Z}(\Delta)$ 就没有必要 (因为这时勒贝格测度几乎处处 $1 - \varphi^{\oplus}(\lambda)\varphi(\lambda) = 0$), 而且 "原概率空间充分 '丰富' " 的前提条件也可以去掉.

系 1 *假设谱密度 (按勒贝格测度几乎处处的) $f_\eta(\lambda) > 0$, 并且*

$$f_\eta(\lambda) = \frac{1}{2\pi}|\varphi(\lambda)|^2,$$

其中

$$\varphi(\lambda) = \sum_{k=0}^{\infty} \mathrm{e}^{-i\lambda k} h(k), \quad \sum_{k=0}^{\infty} |h(k)|^2 < \infty,$$

则序列 η 可以表示为单侧移动平均的形式

$$\eta_n = \sum_{m=0}^{\infty} h(m)\varepsilon_{n-m}.$$

特别, 设 $P(z) = a_0 + a_1 z + \cdots + a_p z^p$, 则谱密度为

$$f_\eta(\lambda) = \frac{1}{2\pi}|P(\mathrm{e}^{-i\lambda})|^2$$

的序列 $\eta = (\eta_n)$ 可以表示为

$$\eta_n = a_0\varepsilon_n + a_1\varepsilon_{n-1} + \cdots + a_p\varepsilon_{n-p}.$$

系 2 假设 $\xi=(\xi_n)$ 是平稳序列, 具有有理谱密度

$$f_{\xi}(\lambda)=\frac{1}{2\pi}\left|\frac{P(\mathrm{e}^{-i\lambda})}{Q(\mathrm{e}^{-i\lambda})}\right|^2, \tag{32}$$

其中 $P(z)=a_0+a_1z+\cdots+a_pz^p, Q(z)=1+b_1z+\cdots+b_qz^q$.

如果多项式 $Q(z)$ 在集合 $\{z:|z|=1\}$ 上无 0 点, 则存在这样的白噪声 $\varepsilon=(\varepsilon_n)$, 使以概率 1, 有

$$\xi_n+b_1\xi_{n-1}+\cdots+b_q\xi_{n-q}=a_0\varepsilon_n+a_1\varepsilon_{n-1}+\cdots+a_p\varepsilon_{n-p}. \tag{33}$$

相反, 设 $\xi=(\xi_n)$ 是任意平稳序列, 并且满足这样带某白噪声 $\varepsilon=(\varepsilon_n)$ 的方程 (33), 以及在集合 $\{z:|z|=1\}$ 上无 0 点的多项式 $Q(z)$, 则 $\xi=(\xi_n)$ 有谱密度 (32) 式.

事实上, 设 $\eta_n=\xi_n+b_1\xi_{n-1}+\cdots+b_q\xi_{n-q}$. 那么,

$$f_{\eta}(\lambda)=\frac{1}{2\pi}|P(\mathrm{e}^{-i\lambda})|^2,$$

而由系 1 得所要求的表达式.

另一方面, 如果表达式 (33) 成立, 而 $F_{\xi}(\lambda)$ 和 $F_{\eta}(\lambda)$ 是序列 ξ 和 η 的谱函数, 则

$$F_{\eta}(\lambda)=\int_{-\pi}^{\lambda}|Q(\mathrm{e}^{-iv})|^2\mathrm{d}F_{\xi}(v)=\frac{1}{2\pi}\int_{-\pi}^{\lambda}|P(\mathrm{e}^{-iv})|^2\mathrm{d}v.$$

由于 $|Q(\mathrm{e}^{-iv})|^2>0$, 故由此得由 (32) 式确定的密度.

4. 遍历性定理 下面 (均方意义下的) *遍历性定理*, 可以看做弱平稳随机序列的大数定律的类似.

定理 4 设 $\xi=(\xi_n), n\in\mathbb{Z}$, 是任意平稳序列, $\mathbf{E}\xi_n=0$, 且具有协方差函数 (1) 式和谱分解 (2) 式, 则

$$\frac{1}{n}\sum_{k=0}^{n-1}\xi_k\xrightarrow{L^2}Z(\{0\}) \tag{34}$$

和

$$\frac{1}{n}\sum_{k=0}^{n-1}R(k)\to F(\{0\}). \tag{35}$$

证明 由 (2) 式, 可见

$$\frac{1}{n}\sum_{k=0}^{n-1}\xi_k=\int_{-\pi}^{\pi}\frac{1}{n}\sum_{k=0}^{n-1}\mathrm{e}^{ik\lambda}Z(\mathrm{d}\lambda)=\int_{-\pi}^{\pi}\varphi_n(\lambda)Z(\mathrm{d}\lambda),$$

其中

$$\varphi_n(\lambda)=\frac{1}{n}\sum_{k=0}^{n-1}\mathrm{e}^{ik\lambda}=\begin{cases}1, & 若\ \lambda=0,\\ \dfrac{1}{n}\times\dfrac{\mathrm{e}^{in\lambda}-1}{\mathrm{e}^{i\lambda}-1}, & 若\ \lambda\neq 0.\end{cases} \tag{36}$$

显然, $|\varphi_n(\lambda)| \leqslant 1$.

其次, 由于 $\varphi_n(\lambda) \xrightarrow{L^2(F)} I_{\{0\}}(\lambda)$, 因此由 §2 的性质 (14), 有

$$\int_{-\pi}^{\pi} \varphi_n(\lambda) Z(\mathrm{d}\lambda) \xrightarrow{L^2} \int_{-\pi}^{\pi} I_{\{0\}}(\lambda) Z(\mathrm{d}\lambda) = Z(\{0\}),$$

于是 (34) 式得证.

类似地可以证明结论 (35) 式. □

系 由于, 如果谱函数在 0 点连续, 即 $F(\{0\}) = 0$, 则 $Z(\{0\}) = 0(\mathbf{P}-\text{a.c.})$, 可见 (34) 式和 (35) 式, 有

$$\frac{1}{n}\sum_{k=0}^{n-1} R(k) \to 0 \Rightarrow \frac{1}{n}\sum_{k=0}^{n-1} \xi_k \xrightarrow{L^2} 0.$$

由于

$$\left|\frac{1}{n}\sum_{k=0}^{n-1} R(k)\right|^2 = \left|\mathbf{E}\left(\frac{1}{n}\sum_{k=0}^{n-1} \xi_k\right)\xi_0\right|^2 \leqslant \mathbf{E}|\xi_0|^2 \mathbf{E}\left|\frac{1}{n}\sum_{k=0}^{n-1} \xi_k\right|^2,$$

可见相反得蕴涵关系也成立:

$$\frac{1}{n}\sum_{k=0}^{n-1} \xi_k \xrightarrow{L^2} 0 \Rightarrow \frac{1}{n}\sum_{k=0}^{n-1} R(k) \to 0.$$

这样, 算术平均值 $n^{-1}\sum\limits_{k=0}^{n-1} \xi_k$ (在均方意义下) 收敛于 0 的充分和必要条件是 $n^{-1}\sum\limits_{k=0}^{n-1} R(k) \to 0$. 由此可见, 如果原序列随机变量的数学期望为 $m(\mathbf{E}\xi_0 = m)$, 则

$$\frac{1}{n}\sum_{k=0}^{n-1} R(k) \to 0 \Leftrightarrow \frac{1}{n}\sum_{k=0}^{n-1} \xi_k \xrightarrow{L^2} 0, \tag{37}$$

其中 $R(n) = \mathbf{E}(\xi_n - \mathbf{E}\xi_n)\overline{(\xi_0 - \mathbf{E}\xi_0)}$.

我们还要指出, 如果以大于 0 的概率 $Z(\{0\}) \neq 0$, 则说明序列 ξ_n 包含 "随机常数 α":

$$\xi_n = \alpha + \eta_n,$$

其中 $\alpha = Z(\{0\})$, 而在谱表示

$$\eta_n = \int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n} Z_\eta(\mathrm{d}\lambda)$$

中测度 $Z_\eta = Z_\eta(\Delta)$ 已经满足 $Z_n(\{0\}) = 0(\mathbf{P}-\text{a.c.})$. 结论 (34) 式说明, 算术平均恰好均方收敛于这一随机常数 α.

5. 练习题

1. 证明 $\overline{L_0^2(F)} = L^2(F)$ (记号见定理 1 的证明).

2. 设 $\xi = (\xi_n)$是平稳序列, 具有性质: 对某个 N 和一切 n, 有 $\xi_{n+N} = \xi_n$. 证明这一序列的谱表示归结为 §1 的表示 (13) 式.

3. 设 $\xi = (\xi_n)$ 是平稳序列, 其中 $\mathbf{E}\xi_n = 0$, 且满足: 对于某 $C > 0, \alpha > 0$, 有

$$\frac{1}{N^2}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1} R(k-l) = \frac{1}{N}\sum_{|k|\leqslant N-1} R(k)\left[1-\frac{|k|}{N}\right] \leqslant \frac{C}{N^\alpha}.$$

利用博雷尔 - 康泰利引理证明

$$\frac{1}{N}\sum_{k=0}^{N}\xi_k \to 0 \quad (\mathbf{P}-\text{a.c.}).$$

4. 假设 $\xi = (\xi_m)$ 是随机变量序列, 具有有理谱密度

$$f_\xi(\lambda) = \frac{1}{2\pi}\frac{|P_{n-1}(\mathrm{e}^{-i\lambda})|}{|Q_n(\mathrm{e}^{-i\lambda})|}, \tag{38}$$

其中 $P_{n-1}(z) = a_0 + a_1 z + \cdots + a_{n-1}z^{n-1}, Q_n(z) = 1 + b_1 z + \cdots + b_n z^n$, 而且多项式 $Q_n(z)$ 的根不在单位圆上.

证明存在这样的白噪声 $\varepsilon = (\varepsilon_m), m \in \mathbb{Z}$, 使序列 (ξ_m) 是 n 维序列 $(\xi_m^1, \xi_m^2, \cdots, \xi_m^n), \xi_m^1 = \xi_m$ 的分量, 而 n 维序列 $(\xi_m^1, \xi_m^2, \cdots, \xi_m^n)$ 满足方程组:

$$\begin{aligned} \xi_{m+1}^i &= \xi_m^{i+1} + \beta_i\varepsilon_{m+1}, \quad i = 1, \cdots, n-1, \\ \xi_{m+1}^n &= -\sum_{j=0}^{n-1} b_{n-j}\xi_m^{j+1} + \beta_n\varepsilon_{m+1}, \end{aligned} \tag{39}$$

其中 $\beta_1 = a_0, \beta_i = a_{i-1} - \sum\limits_{k=1}^{i-1}\beta_k b_{i-k}$.

§4. 协方差函数和谱密度的统计估计

1. 协方差函数的估计及其性质 平稳随机序列概率分布的各种特征的统计估计问题, 出现在各种不同科学领域 (地球物理学, 医学, 经济学等等). 这一节讲述的内容, 建立估计的概念和方法以及这里可能出现的困难.

这样, 设 $\xi = (\xi_n), n \in \mathbb{Z}$, 是广义平稳随机序列 (为简便计, 假设是实数值序列), 且有数学期望 $\mathbf{E}\xi_n = m$, 而其协方差函数有如下表现:

$$R(n) = \int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n}F(\mathrm{d}\lambda).$$

设在对随机变量 $\xi_0,\xi_1,\cdots,\xi_{N-1}$ 值观测过程中得到 (实现) $x_0,x_1\cdots,x_{N-1}$. 问如何根据观测结果建立 (未知) 均值的 "好" 估计?

设

$$m_N(x)=\frac{1}{N}\sum_{k=0}^{N-1}x_k. \tag{1}$$

那么, 由数学期望的初等性质, 可见该估计在如下意义上是数学期望 m 的 "好" 估计: "全部实现 $x_0,\cdots,x_{N-1}$ 的平均值" 是**无偏的**, 即

$$\mathbf{E}m_N(\xi)=\mathbf{E}\left(\frac{1}{N}\sum_{k=0}^{N-1}\xi_k\right)=m. \tag{2}$$

此外, 由 §3 定理 4 可见, 在

$$\frac{1}{N}\sum_{k=0}^{N}R(k)\to 0,\quad N\to\infty,$$

的条件下, 估计在均方意义上也是**相合的**, 即

$$\mathbf{E}|m_N(\xi)-m|^2\to 0,\quad N\to 0. \tag{3}$$

现在, 在 $m=0$ 条件下, 讨论**协方差函数** $R(n)$、谱函数 $F(\lambda)=F([-\pi,\lambda))$ 和**谱密度** $f(\lambda)$ 的估计问题.

由于 $R(n)=\mathbf{E}\xi_{n+k}\xi_k$, 则作为对其根据 N 次观测结果 $x_0,x_1,\cdots,x_{N-1}$ 的估计, 自然假设对于 $0\leqslant n\leqslant N$,

$$\widehat{R}_N(n,x)=\frac{1}{N-n}\sum_{k=0}^{N-n-1}x_{n+k}x_k.$$

显然, 估计在如下意义上是无偏的:

$$\mathbf{E}\widehat{R}_N(n,\xi)=R(n),\quad 0\leqslant n<N.$$

现在讨论估计 $\widehat{R}_N(n,\xi)$ 的**相合性**问题. 将 §3 的 (37) 式中的 ξ_k 换成 $\xi_{n+k}\xi_k$, 并假设对于每一个整数 n, 序列 $\zeta=(\zeta_k)_{k\in\mathbb{Z}},\zeta_k=\xi_{n+k}\xi_k$ 是弱平稳的 (特别, 由此可见存在 4 阶矩 $\mathbf{E}\xi_0^4<\infty$). 因此条件

$$\frac{1}{N}\sum_{k=0}^{N-1}\mathbf{E}[\xi_{n+k}\xi_k-R(n)][\xi_n\xi_0-R(n)]\to 0,\quad N\to\infty, \tag{4}$$

对于

$$\mathbf{E}|\widehat{R}_N(n,\xi)-R(n)|^2\to 0,\quad N\to\infty \tag{5}$$

是充分和必要的.

假设原序列 $\xi=(\xi_n)$ 是 (均值为 0 和协方差为 $R(n)$ 的) 高斯序列, 则由于第二章 §12 的 (51) 式, 有

$$\begin{aligned}&\mathbf{E}[\xi_{n+k}\xi_k-R(n)][\xi_n\xi_0-R(n)]=\mathbf{E}\xi_{n+k}\xi_k\xi_n\xi_0-R^2(n)\\&=\mathbf{E}\xi_{n+k}\xi_k\mathbf{E}\xi_n\xi_0+\mathbf{E}\xi_{n+k}\xi_n\mathbf{E}\xi_k\xi_0+\mathbf{E}\xi_{n+k}\xi_0\mathbf{E}\xi_k\xi_n-R^2(n)\\&=R^2(k)+R(n+k)R(n-k).\end{aligned}$$

因此, 对于高斯情形, 条件 (4) 等价于如下条件:

$$\frac{1}{N}\sum_{k=0}^{N-1}[R^2(k)+R(n+k)R(n-k)]\to 0,\quad N\to\infty. \tag{6}$$

由于 $|R(n+k)R(n-k)|\leqslant|R(n+k)|^2+|R(n-k)|^2$, 故由条件

$$\frac{1}{N}\sum_{k=0}^{N-1}R^2(k)\to 0,\quad N\to\infty, \tag{7}$$

可得到条件 (6). 同时, 若条件 (6) 对于 $n=0$ 成立, 则条件 (7) 也成立.

于是, 我们证明了下面的定理.

定理 *设 $\xi=(\xi_n)$ 是高斯平稳序列, 其均值为 0, 而协方差函数为 $R(n)$. 则条件 (7) 是"对于任何 $n\geqslant 0$, 估计 $\widehat{R}_N(n,\xi)$ 在均方意义下是协方差函数 $R(n)$ 的相合估计"的充分必要条件, 即使"条件 (5) 成立"的充分必要条件.*

注 如果利用协方差函数的谱表达式, 则得

$$\frac{1}{N}\sum_{k=0}^{N-1}R^2(k)=\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{1}{N}\sum_{k=0}^{N-1}\mathrm{e}^{i(\lambda-v)k}F(\mathrm{d}\lambda)F(\mathrm{d}v)=\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}f_N(\lambda,v)F(\mathrm{d}\lambda)F(\mathrm{d}v),$$

其中 (对照 §3 的 (36) 式)

$$f_N(\lambda,v)=\begin{cases}1, & \lambda=v,\\ \dfrac{1-\mathrm{e}^{i(\lambda-v)N}}{N[1-\mathrm{e}^{i(\lambda-v)}]}, & \lambda\neq v.\end{cases}$$

但当 $N\to\infty$

$$f_N(\lambda,v)\to f(\lambda,v)=\begin{cases}1, & \lambda=v,\\ 0, & \lambda\neq v.\end{cases}$$

从而

$$\frac{1}{N}\sum_{k=0}^{N-1}R^2(k)\to\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}f(\lambda,v)F(\mathrm{d}\lambda)F(\mathrm{d}v)=\int_{-\pi}^{\pi}F(\{\lambda\})F(\mathrm{d}\lambda)=\sum_{\lambda}F^2(\{\lambda\}),$$

其中由于测度 F 有限, 故对 λ 的和式中最多有可数项.

于是, 条件 (7) 与条件

$$\sum_{\lambda}F^2(\{\lambda\})=0, \tag{8}$$

等价. 而这说明谱函数 $F(\lambda)=F([-\pi,\lambda])$ 是连续的.

2. 谱函数和谱密度的估计的求法 现在讨论建立谱函数 $F(\lambda)$ 和谱密度 $f(\lambda)$ 的估计问题 (假设 $F(\lambda)$ 和 $f(\lambda)$ 存在).

自然, 由赫尔格洛茨定理 (第 55 页), 可以得出建立谱密度估计必经的途径. 回忆在 §1 引进的函数:

$$f_N(\lambda)=\frac{1}{2\pi}\sum_{|n|<N}\left(1-\frac{|n|}{N}\right)R(n)\mathrm{e}^{-in\lambda},\tag{9}$$

它具有如下性质, 由 $f_N(\lambda)$ 建立的函数

$$F_N(\lambda)=\int_{-\pi}^{\lambda}f_N(v)\mathrm{d}v$$

基本收敛于谱函数 $F(\lambda)$. 因此, 假如谱函数 $F(\lambda)$ 有密度 $f(\lambda)$, 则对于每个 $\lambda\in[-\pi,\pi)$, 有

$$\int_{-\pi}^{\lambda}f_N(v)\mathrm{d}v\to\int_{-\pi}^{\lambda}f(v)\mathrm{d}v.\tag{10}$$

由这些事实, 并注意到 $\widehat{R}_N(n,x)$ 是 (根据观测值 $x_0,x_1,\cdots,x_{N-1}$) 对 $R(n)$ 的估计值, 用函数

$$\widehat{f}_N(\lambda;x)=\frac{1}{2\pi}\sum_{|n|<N}\left(1-\frac{|n|}{N}\right)\widehat{R}_N(n;x)\mathrm{e}^{-i\lambda n},\tag{11}$$

做 $f(\lambda)$ 的估计值, 其中 $\widehat{R}_N(n;x)=\widehat{R}_N(|n|;x)$.

通常称函数 $\widehat{f}_N(\lambda;x)$ 为周期图, 并且不难验证函数 $\widehat{f}_N(\lambda;x)$ 也可以表示为如下较为方便的形式:

$$\widehat{f}_N(\lambda;x)=\frac{1}{2\pi N}\left|\sum_{n=0}^{N-1}x_n\mathrm{e}^{-i\lambda n}\right|^2.\tag{12}$$

由于 $\mathbf{E}\widehat{R}_N(n;\xi)=R(n),|n|<N$, 则

$$\mathbf{E}\widehat{f}_N(\lambda;\xi)=f_N(\lambda).$$

如果谱函数 $F(\lambda)$ 有密度 $f(\lambda)$, 则注意到 $f_N(\lambda)$ 可写成 §1 中 (34) 式的形式, 有

$$\begin{aligned}f_N(\lambda)&=\frac{1}{2\pi N}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}\int_{-\pi}^{\pi}\mathrm{e}^{iv(k-l)}\mathrm{e}^{i\lambda(l-k)}f(v)\mathrm{d}v\\&=\int_{-\pi}^{\pi}\frac{1}{2\pi N}\left|\sum_{k=0}^{N-1}\mathrm{e}^{i(v-\lambda)k}\right|^2f(v)\mathrm{d}v.\end{aligned}$$

函数

$$\Phi_N(\lambda)=\frac{1}{2\pi N}\left|\sum_{k=0}^{N-1}\mathrm{e}^{i\lambda k}\right|^2=\frac{1}{2\pi N}\left|\frac{\sin\dfrac{\lambda}{2}N}{\sin\dfrac{\lambda}{2}}\right|^2$$

称做费耶尔 (L. Féjer) 核. 由该函数的性质知, 对于 (按勒贝格测度) 几乎所有的 λ, 有

$$\int_{-\pi}^{\pi} \Phi_N(\lambda - v) f(v) \mathrm{d}v \to f(\lambda). \tag{13}$$

因此对于几乎所有的 $\lambda \in [-\pi, \pi)$, 有

$$\mathbf{E}\widehat{f}_N(\lambda; \xi) \to f(\lambda), \tag{14}$$

换句话说, 谱函密度 $f(\lambda)$ 根据观测值 $x_0, x_1, \cdots, x_{N-1}$ 的估计 $\widehat{f}_N(\lambda; x)$, 是渐近无偏的.

在此意义上, 可以认为估计 $\widehat{f}_N(\lambda; x)$ 是充分 "好的". 然而, 周期图 $\widehat{f}_N(\lambda; x)$ 在个别观测值 $x_0, x_1, \cdots, x_{N-1}$ 上的值, 对真值 $f(\lambda)$ 的偏离往往是很大的. 例如, 设 $\xi = (\xi_n)$ 是独立平稳高斯随机变量序列, $\xi_n \sim N(0, 1)$. 那么 $f(\lambda) \equiv 1/(2\pi)$, 而

$$\widehat{f}_N(\lambda; \xi) = \frac{1}{2\pi} \left| \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \xi_k \mathrm{e}^{-i\lambda k} \right|^2.$$

因此 $2\pi \widehat{f}_N(0; \xi)$ 与高斯随机变量 $\eta \sim N(0, 1)$ 平方的同分布. 由此可见, 对于任意 N

$$\mathbf{E}|\widehat{f}_N(0; \xi) - f(0)|^2 = \frac{1}{4\pi^2} \mathbf{E}|\eta^2 - 1|^2 > 0.$$

此外, 经不复杂的计算可以证明, 如果 $f(\lambda)$ 是移动平均平稳序列 $\xi = (\xi_n)$:

$$\xi_n = \sum_{k=0}^{\infty} a_k \varepsilon_{n-k}, \tag{15}$$

的谱密度, 而且 $\sum\limits_{k=0}^{\infty} |a_k| < \infty, \sum\limits_{k=0}^{\infty} |a_k|^2 < \infty$, 其中 $\varepsilon = (\varepsilon_n)$ 是白噪声, 且 $\mathbf{E}\varepsilon_0^4 < \infty$, 则

$$\lim_{N \to \infty} \mathbf{E}|\widehat{f}_N(\lambda; \xi) - f(\lambda)|^2 = \begin{cases} 2f^2(0), & \lambda = 0, \pm\pi, \\ f^2(\lambda), & \lambda \neq 0, \pm\pi, \end{cases} \tag{16}$$

由此可见, 周期图不能做谱密度的满意的估计. 为矫正这种情况, 作为 $f(\lambda)$ 的估计常利用形如

$$\widehat{f}_N^W(\lambda; x) = \int_{-\pi}^{\pi} W_N(\lambda - v) \widehat{f}_N(v; x) \mathrm{d}v \tag{17}$$

的估计, 而 $\widehat{f}_N^W(\lambda; x)$ 是由周期图 $\widehat{f}_N(\lambda; x)$ 以及称做谱窗的 "光滑" 函数 $W_N(\lambda)$ 建立的估计. 对于函数 $W_N(\lambda)$ 自然的要求是:

a) $W_N(\lambda)$ 在点 $\lambda = 0$ 的邻域内有 "尖锐的" 极大值;

b) $\int_{-\pi}^{\pi} W_N(\lambda) \mathrm{d}\lambda = 1$;

c) $\mathbf{E}|\widehat{f}_N^W(\lambda; \xi) - f(\lambda)|^2 \to 0, N \to \infty, \lambda \in [-\pi, \pi)$.

由 (14) 式和条件 b) 知, 估计 $\widehat{f}_N^W(\lambda;\xi)$ 是渐近无偏的. 条件 c) 是估计为均方渐近相合性条件, 由上面的讨论知周期图不满足条件 c). 最后, 条件 a) 保障给定频率 λ 的周期图的 "尖锐度".

下面是形如 (17) 式的估计的一些例子.

巴特利特 (M. S. Bartlett) 估计, 基于谱窗

$$W_N(\lambda) = a_N B(a_N \lambda)$$

的选择, 其中 $a_N \uparrow \infty, a_N/N \to 0, N \to \infty$, 而

$$B(\lambda) = \frac{1}{2\pi}\left|\frac{\sin\dfrac{\lambda}{2}}{\dfrac{\lambda}{2}}\right|^2.$$

帕赞 (E. Parzen) 估计, 利用谱窗的函数

$$W_N(\lambda) = a_N P(a_N \lambda),$$

其中 a_N 像巴特利特的情形一样, 而

$$P(\lambda) = \frac{3}{8\pi}\left|\frac{\sin\dfrac{\lambda}{4}}{\dfrac{\lambda}{4}}\right|^4.$$

茹尔边科 (Журбенко) 估计利用形如

$$W_N(\lambda) = a_N Z(a_N \lambda)$$

的谱窗建立的, 其中

$$Z(\lambda) = \begin{cases} -\dfrac{\alpha+1}{2\alpha}|\lambda|^\alpha + \dfrac{\alpha+1}{2\alpha}, & |\lambda| \leqslant 1, \\ 0, & |\lambda| > 1, \end{cases}$$

其中 $0 < \alpha \leqslant 2$, 而 a_N 是专门选择的量.

我们不准备详细讨论谱密度的估计问题, 只是指出, 关于谱窗的建立, 以及与其相应的估计 $\widehat{f}_N^W(\lambda;x)$ 的性质的比较, 存在大量统计学文献. (例如, 见 [133],[71],[72]).

3. 谱函数的估计 现在考虑谱函数 $F(\lambda) = F([-\pi,\lambda])$ 的估计问题. 为此, 设

$$F_N(\lambda) = \int_{-\pi}^{\lambda} f_N(v)\mathrm{d}v, \quad \widehat{F}_N(\lambda;x) = \int_{-\pi}^{\lambda} \widehat{f}_N(v;x)\mathrm{d}v,$$

其中 $\widehat{f}_N(v;x)$ 是根据 $(x_0, x_1, \cdots, x_{N-1})$ 建立的周期图.

由赫尔格洛茨定理 (第 55 页) 的证明, 可见对于任意 $n \in \mathbb{Z}$, 当 $N \to \infty$ 时

$$\int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n}\mathrm{d}F_N(\lambda) \to \int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n}F(\mathrm{d}\lambda).$$

由此 (参照第三章 §3 定理 1 的系) 可得 $F_N \Rightarrow F$, 即在函数 $F(\lambda)$ 的每一个连续点上, $F_N(\lambda)$ 收敛于连续函数 $F(\lambda)$.

注意到, 对于一切 $|n| < N$, 有

$$\int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n} \mathrm{d}\widehat{F}_N(\lambda;\xi) = \widehat{R}_N(n;\xi)\left(1 - \frac{|n|}{N}\right).$$

因此, 如果假设当 $N \to \infty$ 时 $\widehat{R}_N(n;\xi)$ 以概率 1 收敛于 $R(n)$, 则

$$\int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n} \mathrm{d}\widehat{F}_N(\lambda;\xi) \to \int_{-\pi}^{\pi} \mathrm{e}^{i\lambda n} \mathrm{d}\widehat{F}(\lambda) \quad (\mathbf{P}-\text{a.c.}),$$

从而 $\widehat{F}_N(\lambda;\xi) \to F(\lambda) \quad (\mathbf{P}-\text{a.c.})$.

由此容易得到 (当证明必要性时, 需要考虑其子序列), $\widehat{R}_N(n;\xi) \xrightarrow{\mathbf{P}} R(n)$, 而也有 $\widehat{F}_N(\lambda;\xi) \xrightarrow{\mathbf{P}} F$.

4. 练习题

1. 设 (15) 式中的随机变量 $\varepsilon_n \sim N(0,1)$. 证明, 对于任何 n 与当 $N \to \infty$ 时, 有

$$(N - |n|)\mathbf{D}\widehat{R}_N(n;\xi) \to 2\pi \int_{-\pi}^{\pi} (1 + \mathrm{e}^{2in\lambda}) f^2(\lambda)\mathrm{d}\lambda.$$

2. 证明 (16) 式及其如下推广的正确性:

$$\lim_{N\to\infty} \operatorname{cov}(\widehat{f}_N(\lambda;\xi), \widehat{f}_N(v;\xi)) = \begin{cases} 2f^2(0), & \lambda = v = 0, \pm\pi, \\ f^2(\lambda), & \lambda = v \neq 0, \pm\pi, \\ 0, & \lambda \neq v. \end{cases}$$

§5. 沃尔德分解

1. 平稳序列的正则分量和奇异分量 下面将要讨论的沃尔德 (A. Wold) 分解, 与 §3 的表达式 (2) 不同: (2) 式在频率范围内给出了平稳序列的分解, 而沃尔德分解是在时间上的分解. 沃尔德分解的实质在于, 平稳序列 $\xi = (\xi_n), n \in \mathbb{Z}$, 可以表示为两个平稳子列之和: 其中一个可以完全预测 (即它的值完全可以由 "过去" 的值复原), 而另一个却不可以预测.

我们首先引进若干记号. 设 $H_n(\xi) = \overline{L^2}(\xi^n)$ 和 $H(\xi) = \overline{L^2}(\xi)$ 相应为由随机变量 $\xi^n = (\cdots, \xi_{n-1}, \xi_n)$ 和 $\xi = (\cdots, \xi_{n-1}, \xi_n, \cdots)$ 生成的线性流形的闭包. 设

$$S(\xi) = \bigcap_n H_n(\xi).$$

对于任意元素 $\eta \in H(\xi)$, 以

$$\widehat{\pi}_n(\eta) = \widehat{\mathbf{E}}(\eta | H_n(\xi))$$

表示元素 η 在子空间 $H_n(\xi)$ 上的射影 (见第二章 §11). 记

$$\widehat{\pi}_{-\infty}(\eta) = \widehat{\mathbf{E}}[\eta|S(\xi)].$$

每一个元素 $\eta \in H(\xi)$, 可以表示为

$$\eta = \widehat{\pi}_{-\infty}(\eta) + [\eta - \widehat{\pi}_{-\infty}(\eta)],$$

其中 $\eta - \widehat{\pi}_{-\infty}(\eta) \perp \widehat{\pi}_{-\infty}(\eta)$. 因此, 空间 $H(\xi)$ 可以表示为正交和:

$$H(\xi) = S(\xi) \oplus R(\xi),$$

其中 $S(\xi)$ 由形如 $\widehat{\pi}_{-\infty}(\eta)[\eta \in H(\xi)]$ 的元素构成, 而 $R(\xi)$ 由形如 $\eta - \widehat{\pi}_{-\infty}(\eta)$ 的元素构成.

下面, 总是假设 $\mathbf{E}\xi_n = 0, \mathbf{D}\xi_n > 0$. 这样, 空间 $H(\xi)$ 显然是非平凡的 (包含非 0 元素).

定义 1 平稳序列 $\xi = (\xi_n)$ 称做*正则的*, 如果

$$H(\xi) = R(\xi),$$

而称做*奇异的*, 如果

$$H(\xi) = S(\xi).$$

注 1 奇异序列亦称为*确定序列*, 正则序列亦称为*纯粹*或*完全非确定序列*. 如果 $S(\xi)$ 是空间 $H(\xi)$ 的特征空间, 则序列 ξ 称做*非确定序列*.

定理 1 *任意弱平稳随机序列 ξ 可以分解为*

$$\xi_n = \xi_n^r + \xi_n^s, \tag{1}$$

其中 $\xi^r = (\xi_n^r)$ 是正则序列, 而 $\xi^s = (\xi_n^s)$ 是奇异序列. 而且 ξ^r 和 ξ^s 正交 (对于一切 n 和 m, $\xi_n^r \perp \xi_m^s$).

证明 根据定义, 设

$$\xi_n^s = \widehat{\mathbf{E}}[\xi_n|S(\xi)], \quad \xi_n^r = \xi_n - \xi_n^s.$$

由于对于任意 $n, \xi_n^r \perp S(\xi)$, 则 $S(\xi^r) \perp S(\xi)$. 另一方面, 因为 $S(\xi^r) \subseteq S(\xi)$, 说明 $S(\xi^r)$ 退化 (只含几乎必然为 0 的随机变量). 从而, 过程 ξ^r是正则的.

此外, 由于 $H_n(\xi) \subseteq H_n(\xi^s) \oplus H_n(\xi^r)$ 且 $H_n(\xi^s) \subseteq H_n(\xi), H_n(\xi^r) \subseteq H_n(\xi)$, 可见 $H_n(\xi) = H_n(\xi^s) \oplus H_n(\xi^r)$, 因而对于任意 n, 有

$$S(\xi) \subseteq H_n(\xi^s) \oplus H_n(\xi^r). \tag{2}$$

因为 $\xi_n^r \perp S(\xi)$, 所以由 (2) 式, 可见

$$S(\xi) \subseteq H_n(\xi^s),$$

因而 $S(\xi) \subseteq S(\xi^s) \subseteq H(\xi^s)$. 由于 $\xi_n^s \in S(\xi)$, 可见 $H(\xi^s) \subseteq S(\xi)$, 从而

$$S(\xi) = S(\xi^s) = H(\xi^s),$$

这说明序列 ξ^s 是奇异的.

由于 $\xi_n^s \in S(\xi)$, 而且 $\xi_n^r \perp S(\xi)$, 可见显然序列 ξ^r 和 ξ^s 正交. □

注 2　分解 (1) 式中, 将序列 ξ_n 分解为正则分量和奇异分量的唯一性, 见练习题 4.

2. 沃尔德分解

定义 2　设 $\xi = (\xi_n)$ 是非退化平稳序列. 称随机序列 $\varepsilon = (\varepsilon_n)$ (对于 ξ) 为*更新序列*, 如果:

a) $\varepsilon = (\varepsilon_n)$ 由 $\mathbf{E}\varepsilon_n = 0, \mathbf{E}|\varepsilon_n|^2 = 1$ 的两两正交随机变量构成;

b) 对于任意 $n \in \mathbb{Z}, H_n(\xi) = H_n(\varepsilon)$.

注 1　术语 "更新" 的含义与如下联想有关: ε_{n+1} 仿佛携带着 $H_n(\xi)$ 不包含的新的 "信息" (或者说, "更新" $H_n(\xi)$ 中形成 $H_{n+1}(\xi)$ 所必要的信息).

下面的重要定理是前面 (第 52 页例 4) 引进的, 单侧移动平均序列与正则序列之间建立联系.

定理 2　*对于非退化序列 ξ 为正则序列的充分和必要条件是, 存在一更新序列 $\varepsilon = (\varepsilon_n)$ 和复数的序列 $(a_n), n \geqslant 0$, 其中 $\sum\limits_{n=0}^{\infty} |a_n|^2 < \infty$, 使*

$$\xi_n = \sum_{k=0}^{\infty} a_k \varepsilon_{n-k} \quad (\mathbf{P} - \text{a.c.}). \tag{3}$$

证明　(1) *必要性*. 将 $H_n(\xi)$ 表示为:

$$H_n(\xi) = H_{n-1}(\xi) \oplus B_n.$$

由于 $H_n(\xi)$ 是由 $H_{n-1}(\xi)$ 中形如 $\beta\xi_n$ (β 是复数) 的元素组成, 可见空间 B_n 的维数等于 1 或 0. 对于任何 n, $H_n(\xi)$ 都不会与 $H_{n-1}(\xi)$ 等同. 事实上, 假如对于某个 n, B_n 是平凡的, 则由于平稳性, 对于一切 k, B_k 也是平凡的, 而这表示 $H(\xi) = S(\xi)$, 然而这与序列 ξ 的正则性条件相矛盾. 于是, 空间 B_n 的维数等于 1.

假设 η_n 是 B_n 中的非 0 元素, 而

$$\varepsilon_n = \frac{\eta_n}{\|\eta_n\|},$$

其中 $\|\eta_n\|^2 = \mathbf{E}|\eta_n|^2 > 0$.

对于固定的 n 和 k, 考虑分解:

$$H_n(\xi) = H_{n-k}(\xi) \oplus B_{n-k+1} \oplus \cdots \oplus B_n.$$

那么, $\varepsilon_{n-k}, \cdots, \varepsilon_n$ 在 $B_{n-k+1} \oplus \cdots \oplus B_n$ 中构成规范正交基, 且

$$\xi_n = \sum_{j=0}^{k-1} a_j \varepsilon_{n-j} + \widehat{\pi}_{n-k}(\xi_n), \tag{4}$$

其中 $a_j = \mathbf{E}\xi_n \overline{\varepsilon}_{n-j}$.

由贝塞尔不等式 (第二章 §11 的 (6) 式), 可见

$$\sum_{j=0}^{\infty} |a_j|^2 \leqslant \|\xi_n\|^2 < \infty.$$

由此可见, 级数 $\sum\limits_{j=0}^{\infty} a_j \varepsilon_{n-j}$ 在均方意义下收敛, 而由于 (4) 式, 为证明 (3) 式只需证明: 当 $k \to \infty$ 时 $\widehat{\pi}_{n-k}(\xi_n) \xrightarrow{L^2} 0$.

只需证明 $n = 0$ 的情形. 记 $\widehat{\pi}_i = \widehat{\pi}_i(\xi_0)$. 由于

$$\widehat{\pi}_{-k} = \widehat{\pi}_0 + \sum_{i=0}^{k} (\widehat{\pi}_{-i} - \widehat{\pi}_{-i+1}),$$

而和式中的各项正交, 故对于任意 $k \geqslant 0$, 有

$$\sum_{i=0}^{k} \|\widehat{\pi}_{-i} - \widehat{\pi}_{-i+1}\|^2 = \left\|\sum_{i=0}^{k} (\widehat{\pi}_{-i} - \widehat{\pi}_{-i+1})\right\|^2 = \|\widehat{\pi}_{-k} - \widehat{\pi}_0\|^2 \leqslant 4\|\xi_0\|^2 < \infty.$$

于是, 存在均方极限 $\lim\limits_{k\to\infty} \widehat{\pi}_{-k}$. 对于每一个 $k, \widehat{\pi}_{-k} \in H_{-k}(\xi)$, 故所考虑的极限应属于空间 $\widehat{\pi}_{-k}$. 对于每个 $k, \widehat{\pi}_{-k} \in H_{-k}(\xi)$, 所考虑的极限应属于子空间 $\bigcap\limits_{k \geqslant 0} H_{-k}(\xi) = S(\xi)$. 因为根据假设 $S(\xi)$ 是退化的, 所以当 $k \to \infty$ 时 $\widehat{\pi}_{-k} \xrightarrow{L^2} 0$.

(2) *充分性*. 假设非退化序列 ξ 可以表示为 (3) 式的形式, 其中 $\varepsilon = (\varepsilon_n)$ 是正交系 (未必满足条件: $H_n(\xi) = H_n(\varepsilon), n \in \mathbb{Z}$). 那么, $H_n(\xi) \subseteq H_n(\varepsilon)$, 因而对于任意 n, 有 $S(\xi) = \bigcap\limits_k H_k(\xi) \subseteq H_n(\varepsilon)$. 由于 $\varepsilon_{n+1} \perp H_n(\varepsilon)$, 可见 $\varepsilon_{n+1} \perp S(\xi)$, 且 $\varepsilon = (\varepsilon_n)$ 是 $H(\xi)$ 的基. 由此可见, 子空间 $S(\xi)$ 是退化的, 所以序列 ξ 是正则的. □

注 2 由上面的证明, 可见非退化序列 ξ 是*正则的*, 当且仅当它按照 §1 例 4 (第 52 页) 中的定义, 可以表示为*单侧移动平均*的形式:

$$\xi_n = \sum_{k=0}^{\infty} \tilde{a}_k \tilde{\varepsilon}_{n-k}, \tag{5}$$

其中 $\tilde{\varepsilon} = (\tilde{\varepsilon}_n)$ 是某一正交系. 在这种意义上, 由定理 2 的论断可以得到更多的结果. 具体地说, 对于正则序列 ξ, 存在这样的数列 $a = (a_n)$ 和正交系 $\varepsilon = (\varepsilon_n)$, 使与 (5) 式同时 (3) 式也成立, 且对于 (3) 式 $H_n(\xi) = H_n(\varepsilon), n \in \mathbb{Z}$.

由定理 1 和定理 2 直接得下面的定理.

定理 3 (沃尔德分解)　如果 $\xi=(\xi_n)$ 是非退化平稳序列, 则

$$\xi_n=\xi_n^s+\sum_{k=0}^{\infty}a_k\varepsilon_{n-k}, \tag{6}$$

其中 $\sum\limits_{k=0}^{\infty}|a_k|^2<\infty$, 而 $\varepsilon=(\varepsilon_n)$ (对于 ξ^r) 是一更新序列.

3. 外推和预测　在讨论下面的 (线性) 外推问题时, 上面引进的正则和奇异序列的含义, 就显得特别清晰. 利用沃尔德分解 (6) 非常有助于 (线性) 外推问题的一般解.

设 $H_0(\xi)=\overline{L^2}(\xi^0)$ 是由变量 $\xi^0=(\cdots,\xi_{-1},\xi_0)$ 生成的封闭线性流形. 考虑根据 "过去的" 观测结果 $\xi^0=(\cdots,\xi_{-1},\xi_0)$, 建立变量 ξ_n 的 (均方) 最优线性估计 $\widehat{\xi}_n$ 的问题.

由第二章 §11, 可见

$$\widehat{\xi}_n=\widehat{\mathbf{E}}[\xi_n|H_0(\xi)]. \tag{7}$$

(按第 1 小节的记号, $\widehat{\xi}_n=\widehat{\pi}_0(\xi_n)$). 由于 ξ^r 和 ξ^s 正交且 $H_0(\xi)=H_0(\xi^r)\oplus H_0(\xi^s)$, 则注意到 (6) 式, 有

$$\begin{aligned}\widehat{\xi}_n&=\widehat{\mathbf{E}}[\xi_n^s+\xi_n^r|H_0(\xi)]=\widehat{\mathbf{E}}[\xi_n^s|H_0(\xi)]+\widehat{\mathbf{E}}[\xi_n^r|H_0(\xi)]\\&=\widehat{\mathbf{E}}[\xi_n^s|H_0(\xi^r)\oplus H_0(\xi^s)]+\widehat{\mathbf{E}}[\xi_n^r|H_0(\xi^r)\oplus H_0(\xi^s)]\\&=\widehat{\mathbf{E}}[\xi_n^s|H_0(\xi^s)]+\widehat{\mathbf{E}}[\xi_n^r|H_0(\xi^r)]=\xi_n^s+\widehat{\mathbf{E}}\left[\sum_{k=0}^{\infty}a_k\varepsilon_{n-k}\middle|H_0(\xi^r)\right].\end{aligned}$$

在 (6) 式中, 对于 $\varepsilon=(\varepsilon_n),\xi^r=(\xi_n^r)$ 是更新序列, 从而 $H_0(\xi^r)=H_0(\varepsilon)$. 因此

$$\widehat{\xi}_n=\xi_n^s+\widehat{\mathbf{E}}\left[\sum_{k=0}^{\infty}a_k\varepsilon_{n-k}\middle|H_0(\varepsilon)\right]=\xi_n^s+\sum_{k=n}^{\infty}a_k\varepsilon_{n-k}, \tag{8}$$

而根据 $\xi^0=(\cdots,\xi_{-1},\xi_0)$ 预测 ξ_n 的均方误差等于

$$\sigma_n^2=\mathbf{E}|\xi_n-\widehat{\xi}_n|^2=\sum_{k=0}^{n-1}|a_k|^2. \tag{9}$$

由此可得如下两个重要结论:

a) 如果序列 ξ 是奇异的, 则对于任意 $n\geqslant 1$, (外推) 误差 σ_n^2 等于 0. 换句话说, 可以根据 "过去" $\xi^0=(\cdots,\xi_{-1},\xi_0)$ 无误差地预测 ξ_n.

b) 如果序列 ξ 是正则的, 则 $\sigma_n^2\leqslant\sigma_{n+1}^2$, 而且

$$\lim_{n\to\infty}\sigma_n^2=\sum_{k=0}^{\infty}|a_k|^2. \tag{10}$$

由于

$$\sum_{k=0}^{\infty}|a_k|^2=\mathbf{E}|\xi_n|^2,$$

则由 (10) 和 (9) 两式可见

$$\widehat{\xi}_n \xrightarrow{L^2} 0, \quad n\to\infty,$$

即随 n 的增大, 由 $\xi^0=(\cdots,\xi_{-1},\xi_0)$ 预测随机变量 ξ_n 是平凡的 (简单地就等于 $\mathbf{E}\xi_n=0$).

4. 正则序列的充分和必要条件 假设 ξ 是非退化正则平稳序列. 根据定理 2, 任何这样的序列可以表示为单侧移动平均的形式:

$$\xi_n=\sum_{k=0}^{\infty}a_k\varepsilon_{n-k}, \tag{11}$$

其中 $\sum\limits_{k=0}^{\infty}|a_k|^2<\infty$, 规范正交序列 $\varepsilon=(\varepsilon_n)$ 具有如下重要性质:

$$H_n(\xi)=H_n(\varepsilon), \quad n\in\mathbb{Z}. \tag{12}$$

表达式 (11) 表示 (见 §3 第 3 小节), 当在物理可实现的滤波器入口传递序列 $\varepsilon=(\varepsilon_n)$ 时, 以 $a=(a_k), k\geqslant 0$, 为脉冲转移函数的滤波器、在出口的接收信号为 ξ_n.

像任何双侧移动平均一样, 正则序列具有谱密度 $f(\lambda)$. 但是, 由正则序列可以表示为单侧移动平均的情形, 可以得到关于谱密度的补充信息.

首先, 显然

$$f(\lambda)=\frac{1}{2\pi}|\varphi(\lambda)|^2,$$

其中

$$\varphi(\lambda)=\sum_{k=0}^{\infty}\mathrm{e}^{-i\lambda k}a_k, \quad \sum_{k=0}^{\infty}|a_k|^2<\infty. \tag{13}$$

设

$$\Phi(z)=\sum_{k=0}^{\infty}a_kz^k. \tag{14}$$

该函数在开区域 $\{z:|z|<1\}$ 是解析函数, 并且由于条件 $\sum\limits_{n=0}^{\infty}|a_n|^2<\infty$, 属于所谓哈代 (G. H. Hardy) 函数类 H^2, 即在满足

$$\sup_{0\leqslant r<1}\frac{1}{2\pi}\int_{-\pi}^{\pi}|g(r\mathrm{e}^{i\theta})|^2\mathrm{d}\theta<\infty. \tag{15}$$

的、区域 $\{z:|z|<1\}$ 上的解析函数类 $g=g(z)$. 事实上,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}|\Phi(r\mathrm{e}^{i\theta})|^2\mathrm{d}\theta=\sum_{k=0}^{\infty}|a_k|^2r^{2k}$$

且

$$\sup_{0\leqslant r<1}\sum_{k=0}^{\infty}|a_k|^2r^{2k}\leqslant\sum_{k=0}^{\infty}|a_k|^2<\infty.$$

复变函数论中证明, 恒不等于 0 函数 $\Phi\in H^2, -\pi\leqslant\lambda<\pi$ 的、边界值 $\Phi(\mathrm{e}^{i\lambda})$ 满足如下性质:

$$\int_{-\pi}^{\pi}\ln|\Phi(\mathrm{e}^{-i\lambda})|\mathrm{d}\lambda>-\infty. \tag{16}$$

对于现在所考虑的情形,

$$f(\lambda)=\frac{1}{2\pi}|\Phi(\mathrm{e}^{-i\lambda})|^2,$$

其中 $\Phi\in H^2$. 因此

$$\ln f(\lambda)=-\ln 2\pi+2\ln|\Phi(\mathrm{e}^{-i\lambda})|,$$

从而, 正则过程的谱密度 $f(\lambda)$ 满足条件

$$\int_{-\pi}^{\pi}\ln f(\lambda)\mathrm{d}\lambda>-\infty. \tag{17}$$

另一方面, 假设谱密度 $f(\lambda)$ 满足条件 (17). 仍然由复变函数论的性质, 可见在哈代函数类 H^2 中, 存在函数 $\Phi(z)=\sum\limits_{n=0}^{\infty}a_nz^k$, 使 (按勒贝格测度几乎处处)

$$f(\lambda)=\frac{1}{2\pi}|\Phi(\mathrm{e}^{-i\lambda})|^2.$$

因此, 设 $\varphi(\lambda)=\Phi(\mathrm{e}^{-i\lambda})$, 得

$$f(\lambda)=\frac{1}{2\pi}|\varphi(\lambda)|^2,$$

其中 $\varphi(\lambda)$ 是由 (13) 式中的函数. 那么, 由 §3 定理 3 系 1 可见, 序列 ξ 可以表示为单侧移动平均 (11) 的形式, 其中 $\varepsilon=(\varepsilon_n)$ 是某一规范正交序列. 由此以及第 2 小节注 2, 可见序列 ξ 是正则的.

于是, 我们证明了下面的定理.

定理 4 (柯尔莫戈洛夫) *设 ξ 是非退化正则平稳序列, 则存在满足*

$$\int_{-\pi}^{\pi}\ln f(\lambda)\mathrm{d}\lambda>-\infty. \tag{18}$$

谱密度 $f(\lambda)$. 特别, (按勒贝格测度几乎处处) $f(\lambda)>0$.

相反, 如果 ξ 是某一有谱密度且满足条件 (18) 的平稳序列, 则该序列是正则的.

5. 练习题

1. 证明具有离散谱 (谱函数 $F(\lambda)$ 是阶梯函数) 的平稳序列是奇异的.

2. 假设 $\sigma_n^2=\mathbf{E}|\xi_n-\widehat{\xi}_n|^2, \widehat{\xi}_n=\widehat{\mathbf{E}}[\xi_n|H_0(\xi)]$. 证明, 如果对于某个 $n\geqslant1, \sigma_n^2=0$, 则序列 ξ 是奇异的; 而假如 $n\to\infty, \sigma_n^2\to R(0)$, 则序列 ξ 是正则的.

3. 证明平稳序列 $\xi=(\xi_n),\xi_n=\mathrm{e}^{in\varphi}$, 是正则的, 其中 φ 是在 $[0,2\pi]$ 上的均匀分布随机变量. 求估计 $\widehat{\xi}$ 和 σ_n^2 的值. 证明非线性估计

$$\widetilde{\xi}_n=\left(\frac{\xi_0}{\xi_{-1}}\right)^n$$

是根据 "过去" $\xi^0=(\cdots,\xi_{-1},\xi_0)$ 对 ξ_n 的无误差预测, 即

$$\mathbf{E}|\xi_n-\widetilde{\xi}_n|^2=0,\quad n\geqslant 1.$$

4. 证明, 将 ξ_n 分为正则分量和奇异分量的分解 (1) 唯一.

§6. 外推、内插和过滤

1. 外推 对于奇异序列的情形, 由上一节结果, 可以根据 "过去" $\xi^0=(\cdots,\xi_{-1},\xi_0)$, 进行无误差预测 (外推) 随机变量 $\xi_n,n\geqslant 1$. 因此, 自然在讨论任意随机序列的外推问题时, 首先研究正则序列的情形.

根据 §5 的定理 2, 任何正则序列 $\xi=(\xi_n)$ 都可以表示为单侧移动平均

$$\xi_n=\sum_{k=0}^{\infty}a_k\varepsilon_{n-k},\tag{1}$$

的情形, 其中 $\sum\limits_{k=0}^{\infty}|a_k|^2<\infty,\varepsilon=(\varepsilon_n)$ 是某一更新序列. 因为根据 §5 的 (8) 式, 有

$$\widehat{\xi}_n=\sum_{k=n}^{\infty}a_k\varepsilon_{n-k}\tag{2}$$

和

$$\sigma_n^2=\mathbf{E}|\xi_n-\widehat{\xi}_n|^2=\sum_{k=0}^{n-1}|a_k|^2.\tag{3}$$

所以由 §5 可见, 表达式 (1) 解决求最优 (线性) 估计 $\widehat{\xi}_n=\widehat{\mathbf{E}}(\xi_n|H_0(\xi))$ 的问题. 不过, 由于如下的原因, 该解只能是原则性的解.

通常所考虑的序列并没有 (1) 式表达, 而是由其协方差函数 $R(n)$ 或谱密度 $f(\lambda)$ 表达 (对于正则序列 $f(\lambda)$ 存在). 因此, 只有在系数 a_k 通过 $R(n)$ 或 $f(\lambda)$ 值表示, 而变量 ε_k 通过的 $\cdots,\xi_{k-1},\xi_k$ 值表示的情况下, 才可以认为解 (2) 是满意的.

我们不涉及这一问题的一般形式, 而局限于讨论一种 (对应用感兴趣的) 特殊情形, 即谱密度表示为

$$f(\lambda)=\frac{1}{2\pi}|\Phi(\mathrm{e}^{-i\lambda})|^2,\tag{4}$$

的情形, 其中函数 $\Phi(z)=\sum\limits_{k=0}^{\infty}b_kz^k$ 的收敛半径为 $r>1$, 而且在区域 $\{z:|z|\leqslant 1\}$ 内无 0 点.

设

$$\xi_n = \int_{-\pi}^{\pi} e^{i\lambda n} Z(d\lambda) \tag{5}$$

是序列 $\xi = (\xi_n), n \in \mathbb{Z}$, 的谱表示.

定理 1 如果序列 ξ 的谱密度表示为 (4), 则根据 $\xi^0 = (\cdots, \xi_{-1}, \xi_0)$ 对随机变量 ξ_n 的最优 (线性) 估计 $\widehat{\xi}_n$ 由公式

$$\widehat{\xi}_n = \int_{-\pi}^{\pi} \widehat{\varphi}_n(\lambda) Z(d\lambda) \tag{6}$$

表示, 其中

$$\widehat{\varphi}_n(\lambda) = e^{i\lambda n} \frac{\Phi_n(e^{-i\lambda})}{\Phi(e^{-i\lambda})} \tag{7}$$

而

$$\Phi_n(z) = \sum_{k=n}^{\infty} b_k z^k.$$

证明 根据 §3 定理 2 的注, 任意随机变量 $\widetilde{\xi}_n \in H_0(\xi)$ 可以表示为:

$$\widetilde{\xi}_n = \int_{-\pi}^{\pi} \widetilde{\varphi}_n(\lambda) Z(d\lambda), \quad \widetilde{\varphi}_n \in H_0(F), \tag{8}$$

其中 $H_0(F)$ 是由函数

$$e_n = e^{i\lambda n} \ (n \geqslant 0) \left[F(\lambda) = \int_{-\pi}^{\lambda} f(v) dv\right].$$

生成的闭线性流形.

由于

$$\mathbf{E}|\xi_n - \widetilde{\xi}_n|^2 = \left|\mathbf{E}\int_{-\pi}^{\pi} [e^{i\lambda n} - \widetilde{\varphi}_n(\lambda)] Z(d\lambda)\right|^2 = \int_{-\pi}^{\pi} \left|e^{i\lambda n} - \widetilde{\varphi}_n(\lambda)\right|^2 f(\lambda) d\lambda,$$

则估计 (6) 式之最优性的证明归结为证明

$$\inf_{\widetilde{\varphi}_n \in H_0(F)} \int_{-\pi}^{\pi} \left|e^{i\lambda n} - \widetilde{\varphi}_n(\lambda)\right|^2 f(\lambda) d\lambda = \int_{-\pi}^{\pi} \left|e^{i\lambda n} - \widehat{\varphi}_n(\lambda)\right|^2 f(\lambda) d\lambda. \tag{9}$$

由希尔伯特空间的理论 (第二章 §11) 可见, (在 (9) 式意义下的) 最优函数 $\widehat{\varphi}_n(\lambda)$ 决定于如下两个条件:

$$\begin{aligned} &1)\ \widehat{\varphi}_n(\lambda) \in H_0(F); \\ &2)\ [e^{i\lambda n} - \widehat{\varphi}_n(\lambda)] \perp H_0(F). \end{aligned} \tag{10}$$

由于

$$e^{i\lambda n}\Phi_n(e^{-i\lambda}) = e^{i\lambda n}[b_n e^{-i\lambda n} + b_{n+1} e^{-i\lambda(n+1)} + \cdots] \in H_0(F),$$

且类似地 $\Phi_n^{-1}(\mathrm{e}^{-i\lambda}) \in H_0(F)$, 则由 (7) 式定义的函数 $\widehat{\varphi}_n(\lambda)$ 属于函数类 $H_0(F)$. 因此, 为证明函数 $\widehat{\varphi}_n(\lambda)$ 的 "最优性", 只需对于任意 $m \geqslant 0$, 证明

$$[\mathrm{e}^{i\lambda n} - \widehat{\varphi}_n(\lambda)] \perp \mathrm{e}^{-i\lambda m},$$

即

$$I_{n,m} \equiv \int_{-\pi}^{\pi} [\mathrm{e}^{i\lambda n} - \widehat{\varphi}_n(\lambda)]\mathrm{e}^{i\lambda m} f(\lambda)\mathrm{d}\lambda, \quad m \geqslant 0.$$

下面的一系列等式就证明这一论断.

$$\begin{aligned}
I_{n,m} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{i\lambda(n+m)}\left[1 - \frac{\Phi_n(\mathrm{e}^{-i\lambda})}{\Phi(\mathrm{e}^{-i\lambda})}\right] |\Phi(\mathrm{e}^{-i\lambda})|^2 \mathrm{d}\lambda \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{i\lambda(n+m)}[\Phi(\mathrm{e}^{-i\lambda}) - \Phi_n(\mathrm{e}^{-i\lambda})]\overline{\Phi(\mathrm{e}^{-i\lambda})}\mathrm{d}\lambda \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{i\lambda(n+m)}\left(\sum_{k=0}^{n-1} b_k \mathrm{e}^{-i\lambda k}\right)\left(\sum_{l=0}^{\infty} \bar{b}_l \mathrm{e}^{i\lambda l}\right)\mathrm{d}\lambda \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{i\lambda m}\left(\sum_{k=0}^{n-1} b_k \mathrm{e}^{i\lambda(n-k)}\right)\left(\sum_{l=0}^{\infty} \bar{b}_l \mathrm{e}^{i\lambda l}\right)\mathrm{d}\lambda = 0,
\end{aligned}$$

其中最后一等式成立, 因为对于任意 $m \geqslant 0$ 和 $r > 1$,有

$$\int_{-\pi}^{\pi} \mathrm{e}^{i\lambda m}\mathrm{e}^{i\lambda r}\mathrm{d}\lambda = 0. \qquad \square$$

注 1 如果将函数 $\widehat{\varphi}_n(\lambda)$ 展成傅里叶级数

$$\widehat{\varphi}_n(\lambda) = C_0 + C_{-1}\mathrm{e}^{-i\lambda} + C_{-2}\mathrm{e}^{-2i\lambda} + \cdots$$

则根据 "过去" $\xi^0 = (\cdots, \xi_{-1}, \xi_0)$, 对随机变量 $\xi_n, n \geqslant 1$, 的预测 (外推) $\widehat{\xi}_n$ 决定于如下公式:

$$\widehat{\xi}_n = C_0\xi_0 + C_{-1}\xi_{-1} + C_{-2}\xi_{-2} + \cdots$$

注 2 有理函数

$$f(\lambda) = \frac{1}{2\pi}\left|\frac{P(\mathrm{e}^{-i\lambda})}{Q(\mathrm{e}^{-i\lambda})}\right|^2,$$

是由 (4) 式表示的谱密度的典型例子, 其中多项式 $P(z) = a_0 + a_1 z + \cdots + a_p z^p$ 和 $Q(z) = 1 + b_1 z + \cdots + b_q z^q$ 在区域 $\{z : |z| \leqslant 1\}$ 内无 0 点.

事实上, 在这种情形下只需设 $\Phi(z) = P(z)/Q(z)$. 那么

$$\Phi(z) = \sum_{k=0}^{\infty} C_k z^k,$$

而且该级数的收敛半径大于 1.

举两个演示定理 1 的例子.

例 1 设谱密度为

$$f(\lambda)=\frac{1}{2\pi}(5+4\cos\lambda).$$

相应的协方差函数 $R(n)$ 具有“三角形”形状:

$$R(0)=5,\quad R(\pm1)=2,\quad R(n)=0(|n|\geqslant 2). \tag{11}$$

由于所考虑的谱密度可以表示为

$$f(\lambda)=\frac{1}{2\pi}|2+\mathrm{e}^{-i\lambda}|^2,$$

故可以运用定理 1. 易见

$$\widehat{\varphi}_1(\lambda)=\mathrm{e}^{i\lambda}\frac{\mathrm{e}^{-i\lambda}}{2+\mathrm{e}^{-i\lambda}},\quad \widehat{\varphi}_n(\lambda)=0(n\geqslant 2). \tag{12}$$

所以对于一切 $n\geqslant 2, \xi_n=0$, 即根据 $\xi^0=(\cdots,\xi_{-1},\xi_0)$ 对 ξ_n 的 (线性) 预测是平凡的. 因为, 如果注意到根据 (11), $\xi_n(n\geqslant 2)$ 与 $\xi_0,\xi_{-1},\cdots$, 中任意变量的相关性为 0, 则这一点也不奇怪.

对于 $n=1$, 由 (6) 和 (12) 两式, 可见

$$\begin{aligned}\widehat{\xi}_1&=\int_{-\pi}^{\pi}\mathrm{e}^{i\lambda}\frac{\mathrm{e}^{-i\lambda}}{2+\mathrm{e}^{-i\lambda}}Z(\mathrm{d}\lambda)=\frac{1}{2}\int_{-\pi}^{\pi}\frac{1}{1+\dfrac{\mathrm{e}^{-i\lambda}}{2}}Z(\mathrm{d}\lambda)\\&=\sum_{k=0}^{\infty}\frac{(-1)^k}{2^{k+1}}\int_{-\pi}^{\pi}\mathrm{e}^{-ik\lambda}Z(\mathrm{d}\lambda)=\sum_{k=0}^{\infty}\frac{(-1)^k\xi_k}{2^{k+1}}\\&=\frac{1}{2}\xi_0-\frac{1}{4}\xi_{-1}+\cdots\end{aligned}$$

例 2 设协方差函数为

$$R(n)=a^n,\quad |a|<1.$$

那么 (见第 53 页例 5)

$$f(\lambda)=\frac{1}{2\pi}\frac{1-|a|^2}{|1-a\mathrm{e}^{-i\lambda}|^2},$$

即

$$f(\lambda)=\frac{1}{2\pi}|\Phi(\mathrm{e}^{-i\lambda})|^2,$$

其中

$$\Phi(z)=\frac{(1-|a|^2)^{1/2}}{1-az}=(1-|a|^2)^{1/2}\sum_{k=0}^{\infty}(az)^k,$$

由于 $\widehat{\varphi}_n(\lambda)=a^n$, 得

$$\widehat{\xi}_n=\int_{-\pi}^{\pi}a^nZ(\mathrm{d}\lambda)=a^n\xi_0.$$

换句话说, 为由观测值 $\xi^0=(\cdots,\xi_{-1},\xi_0)$ 预测 ξ_n, 只需知道最后一个观测值 ξ_0.

注 3 由正则序列 $\xi=(\xi_n)$ 的沃尔德分解

$$\xi_n=\sum_{k=0}^{\infty}a_k\varepsilon_{n-k}, \tag{13}$$

可见谱密度 $f(\lambda)$ 可以表示为

$$f(\lambda)=\frac{1}{2\pi}|\Phi(\mathrm{e}^{-i\lambda})|^2, \tag{14}$$

其中

$$\Phi(z)=\sum_{k=0}^{\infty}a_kz^k. \tag{15}$$

显然, 相反得命题也成立: 如果 $f(\lambda)$ 可以由形如 (15) 式的函数 $\Phi(z)$ 表示为 (14) 式, 则序列 $\xi=(\xi_n)$ 的沃尔德分解具有 (13) 式的形式. 从而, 谱密度 $f(\lambda)$ 形如 (14) 式的表现问题, 与沃尔德分解中求系数 a_k 的问题等价.

定理 1 中关于函数 $\Phi(z)$ 所作的假设: 对于 $r>1$, 在区域 $\{z:|z|\leqslant 1\}$ 内函数 $\Phi(z)$ 无 0 点, 实际上并不需要. 换句话说, 如果正则序列的谱密度由形如 (14) 的式子表示, 则由 $\xi^0=(\cdots,\xi_{-1},\xi_0)$ 建立的、变量 ξ_n 的均方最优估计由 $\widehat{\xi}_n$ 决定于 (6) 和 (7) 两式.

注 4 定理 1 (连同上面的注 3) 给出了正则序列的预测问题的解. 现在证明, 事实上对于任意平稳序列, 有同样的答案. 确切地说, 设 $\xi_n=\xi_n^s+\xi_n^r$,

$$\xi_n=\int_{-\pi}^{\pi}\mathrm{e}^{i\lambda n}Z(\mathrm{d}\lambda),\quad F(\Delta)=\mathbf{E}|Z(\Delta)|^2,\quad f^r(\lambda)=\frac{1}{2\pi}|\Phi(\mathrm{e}^{-i\lambda})|^2,$$

其中 $f^r(\lambda)$ 是正则序列 $\xi^r=(\xi_n^r)$ 的谱密度. 那么, 估计 $\widehat{\xi}_n$ 决定于 (6) 和 (7) 两式.

实际上 (见 §5 第 3 小节), 设

$$\widehat{\xi}_n=\int_{-\pi}^{\pi}\widehat{\varphi}_n(\lambda)Z(\mathrm{d}\lambda),\quad \widehat{\xi}_n^r=\int_{-\pi}^{\pi}\widehat{\varphi}_n^r(\lambda)Z^r(\mathrm{d}\lambda),$$

其中 $Z^r(\Delta)$ 是正则序列 ξ^r 表现中的正交随机测度. 那么,

$$\begin{aligned}\mathbf{E}|\xi_n-\widehat{\xi}_n|^2&=\int_{-\pi}^{\pi}|\mathrm{e}^{i\lambda n}-\widehat{\varphi}_n(\lambda)|^2F(\mathrm{d}\lambda)\geqslant\int_{-\pi}^{\pi}|\mathrm{e}^{i\lambda n}-\widehat{\varphi}_n(\lambda)|^2f^r(\lambda)\mathrm{d}\lambda\\&\geqslant\int_{-\pi}^{\pi}|\mathrm{e}^{i\lambda n}-\widehat{\varphi}_n^r(\lambda)|^2f^r(\lambda)\mathrm{d}\lambda=\mathbf{E}|\xi_n^r-\widehat{\xi}_n^r|^2.\end{aligned} \tag{16}$$

其次, 由于 $\mathbf{E}|\xi_n-\widehat{\xi}_n|^2=\mathbf{E}|\xi_n^r-\widehat{\xi}_n^r|^2$, 可见 $\xi_n-\widehat{\xi}_n=\xi_n^r-\widehat{\xi}_n^r$, 故由 (16) 式知可以将函数 $\widehat{\varphi}_n(\lambda)$ 选为 $\widehat{\varphi}_n^r(\lambda)$.

2. 内插 假设 $\xi = (\xi_n)$ 是谱密度为 $f(\lambda)$ 的正则序列. 根据对 ξ_0 的 "过去" 值 $\{\xi_n, n = \pm 1, \pm 2, \cdots\}$ 的观测结果, 建立 (均方) 最优线性估计是最简单的内插问题.

以 $H^0(\xi)$ 表示随机变量 $\xi_n (n \neq 0)$ 生成的封闭线性流形. 那么, 根据 §3 的定理 2, 任意随机变量 $\eta \in H^0(\xi)$ 可以表示为:

$$\eta = \int_{-\pi}^{\pi} \varphi(\lambda) Z(\mathrm{d}\lambda),$$

其中 $\varphi \in H^0(F)$ 是函数 $\mathrm{e}^{i\lambda n} (n \neq 0)$ 生成的封闭线性流形, 而估计

$$\check{\xi} = \int_{-\pi}^{\pi} \check{\varphi}(\lambda) Z(\mathrm{d}\lambda) \tag{17}$$

是最优的, 当且仅当

$$\begin{aligned}\inf_{\eta \in H^0(\xi)} \mathbf{E}|\xi_0 - \eta|^2 &= \inf_{\varphi \in H^0(F)} \int_{-\pi}^{\pi} |1 - \varphi(\lambda)|^2 F(\mathrm{d}\lambda) \\ &= \int_{-\pi}^{\pi} |1 - \check{\varphi}_n(\lambda)|^2 F(\mathrm{d}\lambda) = \mathbf{E}|\xi_0 - \check{\xi}_0|^2.\end{aligned}$$

由希尔伯特空间 $H^0(F)$ 中 "垂线" 性质, 可见 (对照 (10) 式) 函数 $\check{\varphi}(\lambda)$ 完全决定于如下两个条件:

$$\begin{aligned}&1)\ \check{\varphi}(\lambda) \in H^0(F); \\ &2)\ 1 - \check{\varphi}(\lambda) \perp H^0(F).\end{aligned} \tag{18}$$

定理 2 (柯尔莫戈洛夫) *设 $\xi = (\xi_n)$ 是谱密度为 $f(\lambda)$ 的正则序列, 且*

$$\int_{-\pi}^{\pi} \frac{\mathrm{d}\lambda}{f(\lambda)} = 0. \tag{19}$$

那么,

$$\check{\varphi}(\lambda) = 1 - \frac{\alpha}{f(\lambda)}, \tag{20}$$

其中

$$\alpha = \frac{2\pi}{\displaystyle\int_{-\pi}^{\pi} \frac{\mathrm{d}\lambda}{f(\lambda)}}, \tag{21}$$

而内插的误差 $\delta^2 = \mathbf{E}|\xi_0 - \check{\xi}_0|^2$ 为 $\delta^2 = 2\pi\alpha$.

证明 证明将在关于谱密度相当严格的条件下进行, 假设

$$0 < c \leqslant f(\lambda) \leqslant C < \infty. \tag{22}$$

由 (18) 式中的条件 2) 可见, 对于任意 $n \neq 0$, 有

$$\int_{-\pi}^{\pi} |1 - \check{\varphi}(\lambda)| \mathrm{e}^{i\lambda n} f(\lambda) \mathrm{d}\lambda = 0. \tag{23}$$

在 (22) 式的条件下, 函数 $|1-\check{\varphi}(\lambda)|f(\lambda)$ 属于带勒贝格测度 $\mathrm{d}\mu$ 的、希尔伯特空间 $L^2([-\pi,\pi],\mathscr{B}([-\pi,\pi]),\mathrm{d}\mu)$. 在此空间中函数系

$$\left\{\frac{\mathrm{e}^{i\lambda n}}{\sqrt{2\pi}},\quad n=0,\pm1,\cdots\right\}$$

规范正交基 (第二章 §12 的练习题 10). 所以由 (23) 式可见, 函数 $|1-\check{\varphi}(\lambda)|f(\lambda)$ 是常数, 记作 α.

这样, 由 (18) 式的第二个条件, 得

$$\check{\varphi}(\lambda)=1-\frac{\alpha}{f(\lambda)}. \tag{24}$$

现在根据 (18) 式的第一个条件, 确定常数 α.

由于 (22) 式可见, $\check{\varphi}\in L^2$ 和 $\check{\varphi}\in H^0(F)$ 与如下的条件等价: $\check{\varphi}$ 属于函数 $\mathrm{e}^{i\lambda n}(n\neq 0)$ 生成的、(在 L^2 中范数的意义上) 封闭线性流形. 由于在函数的分解式中零系数应等于 0, 故

$$0=\int_{-\pi}^{\pi}\check{\varphi}(\lambda)\mathrm{d}\lambda=2\pi-\alpha\int_{-\pi}^{\pi}\frac{\mathrm{d}\lambda}{f(\lambda)},$$

可见常数 α 决定于 (21) 式.

最后,

$$\delta^2=\mathbf{E}|\xi_0-\check{\xi}_0|^2=\int_{-\pi}^{\pi}|1-\check{\varphi}(\lambda)|^2f(\lambda)\mathrm{d}\lambda=|\alpha|^2\int_{-\pi}^{\pi}\frac{f(\lambda)}{f^2(\lambda)}\mathrm{d}\lambda=\frac{4\pi^2}{\displaystyle\int_{-\pi}^{\pi}\frac{\mathrm{d}\lambda}{f(\lambda)}}.$$

于是, 在 (22) 式的补充条件下定理得证. □

系 如果

$$\check{\varphi}(\lambda)=\sum_{0<|k|\leqslant N}c_k\mathrm{e}^{i\lambda k},$$

则

$$\check{\xi}_0=\sum_{0<|k|\leqslant N}c_k\int_{-\pi}^{\pi}\mathrm{e}^{i\lambda k}Z(\mathrm{d}\lambda)=\sum_{0<|k|\leqslant N}c_k\xi_k.$$

例 3 设 $f(\lambda)$ 是上面例 2 中的谱密度, 则通过简单的计算, 可得

$$\check{\xi}_0=\int_{-\pi}^{\pi}\frac{a}{1+|a|^2}(\mathrm{e}^{i\lambda}+\mathrm{e}^{-i\lambda})Z(\mathrm{d}\lambda)=\frac{a}{1+|a|^2}(\xi_1+\xi_{-1}),$$

而内插的误差等于

$$\delta^2=\frac{1-|\alpha|^2}{1+|\alpha|^2}.$$

3. 过滤 设 $(\theta,\xi)=((\theta_n),(\xi_n)), n\in\mathbb{Z}$ 是部分观测序列, 其中 $\theta=(\theta_n)$ 是不被观测的分量, 而 $\xi=(\xi_n)$ 是被观测的分量. 假设 θ 和 ξ 是均值为 0 的 (弱) 平稳序列, 而

$$\theta_n=\int_{-\pi}^{\pi} e^{i\lambda n} Z_\theta(d\lambda), \quad \xi_n=\int_{-\pi}^{\pi} e^{i\lambda n} Z_\xi(d\lambda)$$

相应为 θ 和 ξ 的谱表示. 记

$$F_\theta(\Delta)=\mathbf{E}|Z_\theta(\Delta)|^2, \quad F_\xi(\Delta)=\mathbf{E}|Z_\xi(\Delta)|^2,$$

而

$$F_{\theta\xi}(\Delta)=\mathbf{E}Z_\theta(\Delta)\overline{Z_\xi(\Delta)}.$$

此外, 设 θ 和 ξ 是平稳相联系的, 即其协方差函数 $\operatorname{cov}(\theta_n,\xi_m)=\mathbf{E}\theta_n\overline{\xi}_m$ 只依赖于差 $n-m$, 记 $R_{\theta\xi}(n)=\mathbf{E}\theta_n\overline{\xi}_0$, 则

$$R_{\theta\xi}(n)=\int_{-\pi}^{\pi} e^{i\lambda n} F_{\theta\xi}(d\lambda).$$

所讨论的过滤问题, 在于根据对 ξ 各种不同的观测结果, 建立变量 θ_n 的均方最优线性估计 $\widehat{\theta}_n$.

如果假设是按 $\xi_m, m\in\mathbb{Z}$ 的所有值, 建立变量 θ_n 的估计 $\widehat{\theta}_n$, 则问题的解就特别简单. 事实上, 由于 $\widehat{\theta}_n=\widehat{\mathbf{E}}(\theta_n|H(\xi))$, 则存在这样的函数 $\widehat{\varphi}_n(\lambda)$, 使

$$\widehat{\theta}_n=\int_{-\pi}^{\pi}\widehat{\varphi}_n(\lambda)Z_\xi(d\lambda). \tag{25}$$

如同第 1,2 小节, "最优" 函数 $\widehat{\varphi}_n(\lambda)$ 应该满足的条件如下:

1) $\widehat{\varphi}_n(\lambda)\in H(F_\xi)$;

2) $(\theta_n-\widehat{\theta}_n)\perp H(\xi)$.

由最后一个条件, 可见对于任意 $m\in\mathbb{Z}$, 有

$$\int_{-\pi}^{\pi} e^{i\lambda(n-m)}F_{\theta\xi}(d\lambda)-\int_{-\pi}^{\pi} e^{-i\lambda m}\widehat{\varphi}_n(\lambda)F_\xi(d\lambda)=0. \tag{26}$$

因此, 如果假设函数 $F_{\theta\xi}(\lambda)$ 和 $F_\xi(\lambda)$ 有密度 $f_{\theta\xi}(\lambda)$ 和 $f_\xi(\lambda)$, 则由 (26) 式得

$$\int_{-\pi}^{\pi} e^{i\lambda(n-m)}[f_{\theta\xi}(\lambda)-e^{-i\lambda m}\widehat{\varphi}_n(\lambda)f_\xi(\lambda)]d\lambda=0.$$

如果 (按勒贝格测度几乎处处) $f_\xi(\lambda)>0$, 则由此立即得

$$\widehat{\varphi}_n(\lambda)=e^{i\lambda n}\widehat{\varphi}(\lambda), \tag{27}$$

其中

$$\widehat{\varphi}(\lambda)=f_{\theta\xi}(\lambda)\times f_\xi^{\oplus}(\lambda),$$

而 $f_{\xi}^{\oplus}(\lambda)$ 是 $f_{\xi}(\lambda)$ 的 "广义逆", 即

$$f_{\xi}^{\oplus}(\lambda)=\begin{cases}[f_{\xi}(\lambda)]^{-1}, & f_{\xi}(\lambda)>0,\\ 0, & f_{\xi}(\lambda)=0.\end{cases}$$

这时, 过滤误差

$$\mathbf{E}|\theta_n-\widehat{\theta}_n|^2=\int_{-\pi}^{\pi}[f_{\theta}(\lambda)-f_{\theta\xi}^2(\lambda)f_{\xi}^{\oplus}(\lambda)]\mathrm{d}\lambda. \tag{28}$$

不难验证 $\widehat{\varphi}_n\in H(F_{\xi})$, 从而估计 (25) 式是最优的, 其中函数 $\widehat{\varphi}_n(\lambda)$ 决定于 (27) 式.

例 4 从受到噪声干扰信号中提取信号. 设 $\xi_n=\theta_n+\eta_n$, 其中信号 $\theta=(\theta_n)$ 和噪声 $\eta=(\eta_n)$ 是不相关序列, 其谱密度相应为 $f_{\theta}(\lambda)$ 是 $f_{\eta}(\lambda)$. 那么

$$\widehat{\theta}_n=\int_{-\pi}^{\pi}\mathrm{e}^{i\lambda n}\widehat{\varphi}(\lambda)Z_{\xi}(\mathrm{d}\lambda),$$

其中

$$\widehat{\varphi}(\lambda)=f_{\theta}(\lambda)[f_{\theta}(\lambda)+f_{\eta}(\lambda)]^{\oplus},$$

而过滤误差

$$\mathbf{E}|\theta_n-\widehat{\theta}_n|^2=\int_{-\pi}^{\pi}[f_{\theta}(\lambda)f_{\eta}(\lambda)][f_{\theta}(\lambda)+f_{\eta}(\lambda)]^{\oplus}\mathrm{d}\lambda.$$

现在可以将上面得到的解 (25) 式, 用来由观测结果 $\xi_k(k\leqslant n)$ 建立变量 θ_{n+m} 的最优估计 $\widetilde{\theta}_{n+m}$, 其中 m 是 $\mathbb{Z}$ 中某给定的数. 假设 $\xi=(\xi_n)$ 是正则序列, 其谱密度为

$$f(\lambda)=\frac{1}{2\pi}|\Phi(\mathrm{e}^{-i\lambda})|^2,\quad 其中\quad \Phi(\mathrm{e}^{-i\lambda})=\sum_{k=0}^{\infty}a_k z^k.$$

根据沃尔德分解

$$\xi_n=\sum_{k=0}^{\infty}a_k\varepsilon_{n-k},$$

其中 $\varepsilon=(\varepsilon_n)$ 是白噪声, 其谱分解为

$$\varepsilon_n=\int_{-\pi}^{\pi}\mathrm{e}^{i\lambda n}Z_{\varepsilon}(\mathrm{d}\lambda).$$

因为

$$\widetilde{\theta}_{n+m}=\widehat{\mathbf{E}}[\theta_{n+m}|H_n(\xi)]=\widehat{\mathbf{E}}\left\{\widehat{\mathbf{E}}[\theta_{n+m}|H(\xi)]|H_n(\xi)\right\}=\widehat{\mathbf{E}}[\widehat{\theta}_{n+m}|H_n(\xi)],$$

而

$$\widehat{\theta}_{n+m}=\int_{-\pi}^{\pi}\mathrm{e}^{i\lambda(n+m)}\widehat{\varphi}(\lambda)\Phi(\mathrm{e}^{-i\lambda})Z_{\varepsilon}(\mathrm{d}\lambda)=\sum_{k=-\infty}^{\infty}\widehat{a}_{n+m-k}\varepsilon_k,$$

则

$$\widetilde{\theta}_{n+m}=\widehat{\mathbf{E}}\left[\sum_{k=-\infty}^{\infty}\widehat{a}_{n+m-k}\varepsilon_k|H_n(\xi)\right],$$

其中

$$\widehat{a}_k=\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{e}^{i\lambda k}\widehat{\varphi}(\lambda)\Phi(\mathrm{e}^{-i\lambda})\mathrm{d}\lambda. \tag{29}$$

由于 $H_n(\xi)=H_n(\varepsilon)$, 可见

$$\begin{aligned}\widetilde{\theta}_{n+m}&=\sum_{k\leqslant n}\widehat{a}_{n+m-k}\varepsilon_k=\int_{-\pi}^{\pi}\left[\sum_{k\leqslant n}\widehat{a}_{n+m-k}\mathrm{e}^{i\lambda k}\right]Z_\varepsilon(\mathrm{d}\lambda)\\&=\int_{-\pi}^{\pi}\mathrm{e}^{i\lambda n}\left[\sum_{l=0}^{\infty}\widehat{a}_{l+m}\mathrm{e}^{-i\lambda l}\right]\Phi^{\oplus}(\mathrm{e}^{-i\lambda})Z_\xi(\mathrm{d}\lambda),\end{aligned}$$

其中 $\Phi^{\oplus}$ 是 Φ 的广义逆.

于是, 我们证明了如下定理.

定理 3 *如果被观测的 $\xi=(\xi_n)$ 是正则序列, 则由观测结果 $\xi_k(k\leqslant n)$ 对变量 θ_{n+m} (均方意义下) 的最优线性估计 $\widetilde{\theta}_{n+m}$, 由下面的公式表示:*

$$\widetilde{\theta}_{n+m}=\int_{-\pi}^{\pi}\mathrm{e}^{i\lambda n}H_m(\mathrm{e}^{-i\lambda})Z_\xi(\mathrm{d}\lambda), \tag{30}$$

其中

$$H_m(\mathrm{e}^{-i\lambda})=\sum_{l=0}^{\infty}\widehat{a}_{l+m}\mathrm{e}^{-i\lambda l}\Phi^{\oplus}(\mathrm{e}^{-i\lambda}), \tag{31}$$

而系数 a_k 决定于 (29) 式.

4. 练习题

1. 在定理 1 中, 去掉假设: 级数 $\Phi(z)$ 的收敛半径为 $r>1$, 而且 $\Phi(z)$ 在区域 $\{z:|z|\leqslant 1\}$ 内无 0 点, 并且证明定理 1 仍然成立.

2. 证明, 对于正则过程, (4) 式中的函数 $\Phi(z)$ 可以表示为

$$\Phi(z)=\sqrt{2\pi}\exp\left\{\frac{1}{2}c_0+\sum_{k=1}^{\infty}c_kz^k\right\},\quad |z|<1,$$

其中

$$c_k=\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{e}^{i\lambda k}\ln f(\lambda)\mathrm{d}\lambda.$$

由此导出一步的预测误差 $\sigma_1^2=\mathbf{E}|\widehat{\xi}_1-\xi_1|^2$, 由塞格 — 柯尔莫戈洛夫 (G. Szegö – A. H. Колмогоров) 公式

$$\sigma_1^2=2\pi\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln f(\lambda)\mathrm{d}\lambda\right\}$$

给出.

3. 不要求条件 (22) 成立, 证明定理 2.

4. 设不相关信号 θ 和噪声 η 的谱密度分别为;

$$f_\theta(\lambda)=\frac{1}{2\pi}\times\frac{1}{|1+b_1\mathrm{e}^{-i\lambda}|^2},\quad f_\eta(\lambda)=\frac{1}{2\pi}\times\frac{1}{|1+b_2\mathrm{e}^{-i\lambda}|^2}.$$

基于定理 3, 根据 $\xi_k(k\leqslant n)$ 的值求量 θ_{n+m} 的估计 $\widetilde{\theta}_{n+m}$, 其中 $\xi_k=\theta_k+\eta_k$. 对于谱密度

$$f_\theta(\lambda)=\frac{1}{2\pi}\times|2+\mathrm{e}^{-i\lambda}|^2,\quad f_\eta(\lambda)=\frac{1}{2\pi},$$

讨论同一问题.

§7. 卡尔曼 – 布西滤波器及其推广

1. 卡尔曼 – 布西概型, 卡尔曼 – 布西滤波器 从计算的角度, 上面给出的由对 ξ 的观测, 求解非观测分量 θ 的过滤问题并不适宜, 因为既然用谱的语言表述, 它对于自己的实现就要求进行类似的处理. 在卡尔曼 – 布西 (R. E. Kalman–R. S. Bucy) 提出的概型中, 最优滤波器的综合是用递推的方法实现的, 从而可能借助数字计算设备来实现. 决定卡尔曼 – 布西滤波器的广泛应用还有其他原因. 在不假设 (θ,ξ) 是平稳序列的情况下, 卡尔曼 – 布西滤波器仍然能 “运用” 就是其有广泛应用的原因之一.

我们在下面不仅讨论传统的卡尔曼 – 布西概型, 而且也研究其推广: 决定 (θ,ξ) 的递推方程中系数可能依赖于全部以往观测数据的情形.

这样, 假设 $(\theta,\xi)=((\theta_n),(\xi_n))$ 是部分可观测序列, 并且

$$\theta_n=(\theta_1(n),\cdots,\theta_k(n)),\quad \xi_n=(\xi_1(n),\cdots,\xi_l(n))$$

满足递推方程

$$\begin{aligned}\theta_{n+1}&=a_0(n,\xi)+a_1(n,\xi)\theta_n+b_1(n,\xi)\varepsilon_1(n+1)+b_2(n,\xi)\varepsilon_2(n+1),\\ \xi_{n+1}&=A_0(n,\xi)+A_1(n,\xi)\theta_n+B_1(n,\xi)\varepsilon_1(n+1)+B_2(n,\xi)\varepsilon_2(n+1),\end{aligned}\tag{1}$$

其中 $\varepsilon_1(n)=(\varepsilon_{11}(n),\cdots,\varepsilon_{1k}(n)),\varepsilon_2(n)=(\varepsilon_{21}(n),\cdots,\varepsilon_{2l}(n))$ 是具有独立分量的相互独立的高斯向量, 并且每一个分量服从参数 0 和 1 的高斯分布; $a_0(n,\xi)=(a_{01}(n,\xi),\cdots,a_{0k}(n,\xi))$ 和 $A_0(n,\xi)=(A_{01}(n,\xi),\cdots,A_{0l}(n,\xi))$ 是向量函数, 而且这些向量函数与 $\xi=(\xi_0,\xi_1,\cdots)$ 的独立性是 “不超前的”, 也就是说, 对于任意固定的 $n,a_{01}(n,\xi),\cdots,A_{0l}(n,\xi)$, 仅依赖于$\xi_0,\cdots,\xi_n$. 矩阵函数

$$\begin{aligned}b_1(n,\xi)&=\|b_{ij}^{(1)}(n,\xi)\|,&b_2(n,\xi)&=\|b_{ij}^{(2)}(n,\xi)\|,\\ B_1(n,\xi)&=\|B_{ij}^{(1)}(n,\xi)\|,&B_2(n,\xi)&=\|B_{ij}^{(2)}(n,\xi)\|,\\ a_1(n,\xi)&=\|a_{ij}^{(1)}(n,\xi)\|,&A_1(n,\xi)&=\|A_{ij}^{(1)}(n,\xi)\|\end{aligned}$$

的阶数相应地为 $k\times k, k\times l, l\times k, l\times l, k\times k, l\times k$, 并且亦 "不超前" 地依赖于 ξ. 此外, 假设向量 (θ_0,ξ_0) 不依赖于序列 $\varepsilon_1=(\varepsilon_1(n))$ 和 $\varepsilon_2=(\varepsilon_2(n))$.

为叙述简便, 以下将省略 "系数不依赖于 ξ" 的提示.

为使方程组 (1) 有解, 假设

$$\mathbf{E}(\|\theta_0\|^2+\|\xi_0\|^2)<\infty, \quad \left(\|x\|^2=\sum_{i=1}^{k}x_i^2, x=(x_1,\cdots,x_k)\right),$$

$|a_{ij}^{(1)}(n,\xi)|\leqslant C, A_{ij}^{(1)}(n,\xi)\leqslant C$, 而且如果 $g(n,\varepsilon)$ 是 $a_{0i}, A_{0j}, b_{ij}^{(1)}, b_{ij}^{(2)}, B_{ij}^{(1)}, B_{ij}^{(2)}$ 任意函数, 则 $\mathbf{E}|g(n,\xi)|^2<\infty, n=1,2,\cdots$. 在这些条件下, 对于序列 (θ,ξ) 满足 $\mathbf{E}(\|\theta_n\|^2+\|\xi_n\|^2)<\infty, n\geqslant 1$.

其次, 设 $\mathscr{F}_n^\xi=\sigma\{\omega:\xi_0,\cdots,\xi_n\}$ 是随机变量 $\xi_0,\cdots,\xi_n$ 生成的最小 $\sigma-$ 代数, 并且

$$m_n=\mathbf{E}(\theta_n|\mathscr{F}_n^\xi), \quad \gamma_n=\mathbf{E}[(\theta_n-m_n)(\theta_n-m_n)^*|\mathscr{F}_n^\xi].$$

根据第二章 §8 定理 1, $m_n=(m_1(n),\cdots,m_k(n))$ 是向量 $\theta_n=(\theta_1(n),\cdots,\theta_k(n))$ 的均方最优估计量, 而 $\mathbf{E}\gamma_n=\mathbf{E}[(\theta_n-m_n)(\theta_n-m_n)^*]$ 是估计误差矩阵. 对于由方程 (1) 决定的任意序列 (θ,ξ), 求这些变量是相当困难的课题. 不过, 在如下关于 (θ_0,ξ_0) 一个补充条件下, 就可以导出 m_n 和 γ_n 递推方程组, 其中包含所谓*卡尔曼 — 布西滤波方程*. 关于 (θ_0,ξ_0) 的这个补充假设是: 条件分布 $\mathbf{P}\{\theta_0\leqslant a|\xi_0\}$ 是参数为高斯分布

$$\mathbf{P}\{\theta_0\leqslant a|\xi_0\}=\frac{1}{\sqrt{2\pi\gamma_0}}\int_{-\infty}^{a}\mathrm{e}^{-\frac{(x-m_0)^2}{2\gamma_0^2}}\mathrm{d}x, \tag{2}$$

其中 $m_0=m_0(\xi_0), \gamma_0=\gamma_0(\xi_0)$ 是分布参数.

首先证明一个辅助命题.

引理 1 *在上面引进的关于系数方程组 (1) 和条件 (2) 补充假设下, 序列 (θ,ξ) 服从条件高斯分布, 即条件分布函数*

$$\mathbf{P}\{\theta_0\leqslant a_0,\cdots,\theta_n\leqslant a_n|\mathscr{F}_n^\xi\}$$

以概率 1 是 n 维高斯向量的分布函数, 其分布的均值和协方差矩阵都依赖于 $\xi_0,\cdots,\xi_n$.

证明 我们仅限于证明分布 $\mathbf{P}\{\theta_n\leqslant a_n|\mathscr{F}_n^\xi\}$ 的高斯性, 因此只需导出 m_n 和 γ_n 的方程.

首先注意到, 由 (1) 式可见条件分布

$$\mathbf{P}\{\theta_{n+1}\leqslant a, \xi_{n+1}\leqslant x|\mathscr{F}_n^\xi, \theta_n=b\}$$

是高斯的, 其均值向量为

$$\mathbb{A}_0+\mathbb{A}_1 b=\begin{pmatrix}a_0+a_1b\\A_0+A_1b\end{pmatrix},$$

而方差矩阵为

$$\mathbb{B} = \begin{pmatrix} b \circ b & b \circ B \\ (b \circ B)^* & (B \circ B) \end{pmatrix},$$

其中 $b \circ b = b_1 b_1^* + b_2 b_2^*, b \circ B = b_1 B_1^* + b_2 B_2^*, B \circ B = B_1 B_1^* + B_2 B_2^*$.

记 $\zeta_n = (\theta_n, \xi_n)$ 和 $t = (t_1, \cdots, t_{k+l})$, 则

$$\begin{aligned} &\mathbf{E}[\exp\{it^* \zeta_{n+1}\} | \mathscr{F}_n^\xi, \theta_n] \\ &= \exp\left\{ it^*[\mathbb{A}_0(n,\xi) + \mathbb{A}_1(n,\xi)\theta_n] - \frac{1}{2} t^* \mathbb{B}(n,\xi) t \right\}. \end{aligned} \tag{3}$$

现在, 假设对于某各 $n \geqslant 0$, 引理的结论成立. 那么,

$$\begin{aligned} &\mathbf{E}[\exp\{it^* \mathbb{A}_1(n,\xi)\theta_n\} | \mathscr{F}_n^\xi] \\ &= \exp\left\{ it^* \mathbb{A}_1(n,\xi) m_n - \frac{1}{2} t^* [\mathbb{A}_1(n,\xi)\gamma_n \mathbb{A}_1^*(n,\xi)] t \right\}. \end{aligned} \tag{4}$$

证明当将 n 换成 $n+1$ 时 (4) 式仍然成立.

由 (3) 式和 (4) 式, 有

$$\begin{aligned} &\mathbf{E}[\exp\{it^* \zeta_{n+1}\} | \mathscr{F}_n^\xi] \\ &= \exp\left\{ it^* [\mathbb{A}_0(n,\xi) + \mathbb{A}_1(n,\xi) m_n] - \frac{1}{2} t^* \mathbb{B}(n,\xi) t - \frac{1}{2} t^* [\mathbb{A}_1(n,\xi)\gamma_n \mathbb{A}_1^*(n,\xi)] t \right\}. \end{aligned}$$

因此, 条件分布

$$\mathbf{P}\{\theta_{n+1} \leqslant a, \xi_{n+1} \leqslant x | \mathscr{F}_n^\xi\} \tag{5}$$

是高斯的.

像证明正态相关性定理 (第二章 §13 定理 2) 时一样, 验证存在矩阵 C, 使向量

$$\eta = [\theta_{n+1} - \mathbf{E}(\theta_{n+1} | \mathscr{F}_n^\xi)] - C[\xi_{n+1} - \mathbf{E}(\xi_{n+1} | \mathscr{F}_n^\xi)]$$

以概率 1 具有性质:

$$\mathbf{E}\left\{ \eta [\xi_{n+1} - \mathbf{E}(\xi_{n+1} | \mathscr{F}_n^\xi)]^* | \mathscr{F}_n^\xi \right\} = 0.$$

由此可见, 在 $\mathscr{F}_n^\xi$ 的条件下, 条件高斯向量 η 和 ξ_{n+1} 相互独立, 即对于任意 $A \in \mathscr{B}(\mathbb{R}^k)$和 $B \in \mathscr{B}(\mathbb{R}^l)$, 以概率 1, 有

$$\mathbf{P}\{\eta \in A, \xi_{n+1} \in B | \mathscr{F}_n^\xi\} = \mathbf{P}\{\eta \in A | \mathscr{F}_n^\xi\} \times \mathbf{P}\{\xi_{n+1} \in B | \mathscr{F}_n^\xi\}.$$

因此, 对于任意 $s = (s_1, \cdots, s_k)$, 有

$$\begin{aligned} &\mathbf{E}[\exp(is^* \theta_{n+1}) | \mathscr{F}_n^\xi, \xi_{n+1}] \\ &= \mathbf{E}\left\{ \exp\left(is^* \left[\mathbf{E}(\theta_{n+1} | \mathscr{F}_n^\xi) + \eta + C[\xi_{n+1} - \mathbf{E}(\xi_{n+1} | \mathscr{F}_n^\xi)] \right] \right) \Big| \mathscr{F}_n^\xi, \xi_{n+1} \right\} \\ &= \exp\left\{ is^* \left[\mathbf{E}(\theta_{n+1} | \mathscr{F}_n^\xi) + C[\xi_{n+1} - \mathbf{E}(\xi_{n+1} | \mathscr{F}_n^\xi)] \right] \right\} \mathbf{E}[\exp(is^* \eta) | \mathscr{F}_n^\xi, \xi_{n+1}] \\ &= \exp\left\{ is^* \left[\mathbf{E}(\theta_{n+1} | \mathscr{F}_n^\xi) + C[\xi_{n+1} - \mathbf{E}(\xi_{n+1} | \mathscr{F}_n^\xi)] \right] \right\} \mathbf{E}[\exp(is^* \eta) | \mathscr{F}_n^\xi]. \end{aligned} \tag{6}$$

根据 (5) 式, 条件分布函数 $\mathbf{P}\{\eta \leqslant y|\mathscr{F}^{\xi}_{n}\}$是高斯的. 这连同 (6) 式, 就证明了条件分布 $\mathbf{P}\{\theta_{n+1} \leqslant a|\mathscr{F}^{\xi}_{n+1}\}$ 也是高斯的. □

定理 1 设 (θ,ξ) 是满足 (1) 式和 (2) 式的部分观测序列. 那么, (m_n,γ_n) 服从下列递推方程:

$$\begin{aligned} m_{n+1} = &[a_0 + a_1 m_n] \\ &+[b \circ B + a_1\gamma_n A_1^*][B \circ B + A_1\gamma_n A_1^*]^{\oplus}[\xi_{n+1} - A_0 - A_1 m_n], \end{aligned} \tag{7}$$

$$\begin{aligned} \gamma_{n+1} = &[a_1\gamma_n a_1^* + b \circ b] \\ &-[b \circ B + a_1\gamma_n A_1^*][B \circ B + A_1\gamma_n A_1^*]^{\oplus}[b \circ B + a_1\gamma_n A_1^*]^*. \end{aligned} \tag{8}$$

证明 由 (1) 式, 有

$$\mathbf{E}(\theta_{n+1}|\mathscr{F}^{\xi}_{n}) = a_0 + a_1 m_n, \quad \mathbf{E}(\xi_{n+1}|\mathscr{F}^{\xi}_{n}) = A_0 + A_1 m_n \tag{9}$$

和

$$\begin{aligned} \theta_{n+1} - \mathbf{E}(\theta_{n+1}|\mathscr{F}^{\xi}_{n}) &= a_1(\theta_n - m_n) + b_1\varepsilon_1(n+1) + b_2\varepsilon_2(n+1), \\ \xi_{n+1} - \mathbf{E}(\xi_{n+1}|\mathscr{F}^{\xi}_{n}) &= A_1(\theta_n - m_n) + B_1\varepsilon_1(n+1) + B_2\varepsilon_2(n+1). \end{aligned} \tag{10}$$

引进记号:

$$\begin{aligned} d_{11} &= \operatorname{cov}(\theta_{n+1},\theta_{n+1}|\mathscr{F}^{\xi}_{n}) \\ &= \mathbf{E}\{[\theta_{n+1} - \mathbf{E}(\theta_{n+1}|\mathscr{F}^{\xi}_{n})][\theta_{n+1} - \mathbf{E}(\theta_{n+1}|\mathscr{F}^{\xi}_{n})]^*|\mathscr{F}^{\xi}_{n}\}, \\ d_{12} &= \operatorname{cov}(\theta_{n+1},\xi_{n+1}|\mathscr{F}^{\xi}_{n}) \\ &= \mathbf{E}\{[\theta_{n+1} - \mathbf{E}(\theta_{n+1}|\mathscr{F}^{\xi}_{n})][\xi_{n+1} - \mathbf{E}(\xi_{n+1}|\mathscr{F}^{\xi}_{n})]^*|\mathscr{F}^{\xi}_{n}\}, \\ d_{22} &= \operatorname{cov}(\xi_{n+1},\xi_{n+1}|\mathscr{F}^{\xi}_{n}) \\ &= \mathbf{E}\{[\xi_{n+1} - \mathbf{E}(\xi_{n+1}|\mathscr{F}^{\xi}_{n})][\xi_{n+1} - \mathbf{E}(\xi_{n+1}|\mathscr{F}^{\xi}_{n})]^*|\mathscr{F}^{\xi}_{n}\}. \end{aligned}$$

那么, 由 (10) 式, 得

$$d_{11} = a_1\gamma_n a_1^* + b \circ b, \quad d_{12} = a_1\gamma_n A_1^* + b \circ B, \quad d_{22} = A_1\gamma_n A_1^* + B \circ B. \tag{11}$$

由正态相关定理 (见第二章 §13 定理 2 和练习题 4), 有

$$m_{n+1} = \mathbf{E}(\theta_{n+1}|\mathscr{F}^{\xi}_{n},\xi_{n+1}) = \mathbf{E}(\theta_{n+1}|\mathscr{F}^{\xi}_{n}) + d_{12}d_{22}^{\oplus}[\xi_{n+1} - \mathbf{E}(\xi_{n+1}|\mathscr{F}^{\xi}_{n})]$$

和

$$\gamma_{n+1} = \operatorname{cov}(\theta_{n+1},\theta_{n+1}|\mathscr{F}^{\xi}_{n},\xi_{n+1}) = d_{11} - d_{12}d_{22}^{\oplus}d_{12}^*.$$

将 (9) 式中的 $\mathbf{E}(\theta_{n+1}|\mathscr{F}^{\xi}_{n})$ 和 $\mathbf{E}(\xi_{n+1}|\mathscr{F}^{\xi}_{n})$, 代入上面 m_{n+1} 的表达式, 将 (11) 式中的 d_{11},d_{12},d_{22}, 代入上面 γ_{n+1} 的表达式, 即可得到欲证明的递推方程 (7) 和 (8). □

系 1 如果在方程组 (1) 中, 所有系数 $a_0(n,\xi),\cdots,B_2(n,\xi)$ 都不依赖于 ξ, 则相应的概型称做卡尔曼－布西概型, 而关于 m_n 和 γ_n 的方程 (7) 和 (8) 称做卡尔曼－布西滤波器. 需要强调, 在这种情况下, 条件误差矩阵 γ_n 的等于无条件矩阵, 即

$$\gamma_n \equiv \mathbf{E}\gamma_n = \mathbf{E}[(\theta_n - m_n)(\theta_n - m_n)^*].$$

系 2 假设对于部分观测序列 (θ_n,ξ_n), θ_n 满足 (1) 式的第一个方程, 而 ξ_n 满足方程:

$$\begin{aligned}\xi_n = \widetilde{A}_0(n-1,\xi) &+ \widetilde{A}_1(n-1,\xi)\theta_n \\ &+ \widetilde{B}_1(n-1,\xi)\varepsilon_1(n) + \widetilde{B}_2(n-1,\xi)\varepsilon_2(n).\end{aligned} \tag{12}$$

那么, 显然

$$\begin{aligned}\xi_{n+1} = \widetilde{A}_0(n,\xi) + \widetilde{A}_1(n,\xi)[a_0(n,\xi) + a_1(n,\xi)\theta_n + b_1(n,\xi)\varepsilon_1(n+1) \\ + b_2(n,\xi)\varepsilon_2(n+1)] + \widetilde{B}_1(n,\xi)\varepsilon_1(n+1) + \widetilde{B}_2(n,\xi)\varepsilon_2(n+1).\end{aligned}$$

若记

$$\begin{aligned} A_0 &= \widetilde{A}_0 + \widetilde{A}_1 a_0, \quad & A_1 &= \widetilde{A}_1 a_1, \\ B_1 &= \widetilde{A}_1 b_1 + \widetilde{B}_1, \quad & B_2 &= \widetilde{A}_2 b_2 + \widetilde{B}_2,\end{aligned}$$

则所考虑的情形也属于概型 (1), 而且 m_n 和 γ_n 满足方程 (7) 和 (8).

2. 最优线性滤波器的结构 现在考虑线性概型 (对照 (1)式)

$$\begin{aligned}\theta_{n+1} &= a_0 + a_1\theta_n + a_2\xi_n + b_1\varepsilon_1(n+1) + b_2\varepsilon_2(n+1), \\ \xi_{n+1} &= A_0 + A_1\theta_n + A_2\xi_n + B_1\varepsilon_1(n+1) + B_2\varepsilon_2(n+1),\end{aligned} \tag{13}$$

其中所有系数 $a_0,\cdots,B_2$ 可能依赖于 n (但不依赖于 ξ), 而且 $\varepsilon_{ij}(n)$ 是 $\mathbf{E}\varepsilon_{ij}(n)=0$ 和 $\mathbf{E}\varepsilon_{ij}^2(n)=1$ 的独立高斯随机变量.

假设方程组 (13) 在初始条件 (θ_0,ξ_0) 下求解, 而且条件分布 $\mathbf{P}\{\theta_0 \leqslant a|\xi_0\}$ 是参数为 $m_0=\mathbf{E}(\theta_0|\xi_0)$ 和 $\gamma_0=\operatorname{cov}(\theta_0,\theta_0|\xi_0)=\mathbf{E}\gamma_0$ 的高斯分布. 那么, 由于正态相关定理以及 (7) 和 (8) 式, $m_n=\mathbf{E}(\theta_n|\mathscr{F}_n^{\xi})$ 的最优估计量是 $\xi_0,\xi_1,\cdots,\xi_n$ 的线性函数.

由以上讨论的结果, 在不要求高斯性的情形下, 可以证明关于 "线性滤波器结构" 的如下重要论断.

定理 2 设 $(\theta,\xi)=(\theta_n,\xi_n)_{n\geqslant 0}$ 是满足方程组 (13) 的部分被观测序列, 其中 $\varepsilon_{ij}(n)$ 是不相关随机变量, 且 $\mathbf{E}\varepsilon_{ij}(n)=0$ 和 $\mathbf{E}\varepsilon_{ij}^2(n)=1$, 而初值向量 (θ_0,ξ_0) 的分量具有有限二阶矩. 那么, $m_n=\mathbf{E}(\theta_n|\mathscr{F}_n^{\xi})$ 的最优线性估计量 $\widehat{m}_n=\widehat{\mathbf{E}}(\theta_n|\xi_0,\xi_1,\cdots,\xi_n)$ 满足方程 (7), 其中

$$a_0(n,\xi)=a_0(n)+a_2(n)\xi_n, \quad A_0(n,\xi)=A_0(n)+A_2(n)\xi_n,$$

而误差矩阵 $\widehat{\gamma} = \widehat{\mathbf{E}}[(\theta_n - \widehat{m}_n)(\theta_n - \widehat{m}_n)^*]$ 决定于方程 (8), 其初始条件为

$$\begin{aligned}\widehat{m}_0 &= \operatorname{cov}(\theta_0, \xi_0)\operatorname{cov}^{\oplus}(\xi_0, \xi_0)\xi_0, \\ \widehat{\gamma}_0 &= \operatorname{cov}(\theta_0, \theta_0) - \operatorname{cov}(\theta_0, \xi_0)\operatorname{cov}^{\oplus}(\xi_0, \xi_0)\operatorname{cov}^*(\theta_0, \xi_0).\end{aligned} \tag{14}$$

定理的证明需要下面的引理. 在建立最优线性估计时, 引理说明高斯性作用.

引理 2 设 (α, β) 是二维随机向量, 且 $\mathbf{E}(\alpha^2 + \beta^2) < \infty$, 而 $(\widetilde{\alpha}, \widetilde{\beta})$ 是与 (α, β) 具有相同一阶和二阶矩的二维高斯向量, 即

$$\mathbf{E}\widetilde{\alpha}^i = \mathbf{E}\alpha^i, \quad \mathbf{E}\widetilde{\beta}^i = \mathbf{E}\beta^i, \quad i = 1, 2, \quad \mathbf{E}\widetilde{\alpha}\widetilde{\beta} = \mathbf{E}\alpha\beta.$$

设 $\lambda(b)$ 是 b 的线性函数, 且

$$\lambda(b) = \mathbf{E}(\widetilde{\alpha}|\widetilde{\beta} = b).$$

那么, $\lambda(\beta)$ 是由 β 对 α 的均方最优线性估计, 即

$$\widehat{\mathbf{E}}(\alpha|\beta) = \lambda(\beta).$$

这时 $\mathbf{E}\lambda(\beta) = \mathbf{E}\alpha$.

证明 首先注意到, 由正态相关定理, 可见存在线性函数 $\lambda(b)$: $\lambda(b) = \mathbf{E}(\widetilde{\alpha}|\widetilde{\beta} = b)$. 其次, 假设 $\overline{\lambda}(b)$ 是另外一个线性估计, 那么

$$\mathbf{E}[\widetilde{\alpha} - \overline{\lambda}(\beta)]^2 \geqslant \mathbf{E}[\widetilde{\alpha} - \lambda(\widetilde{\beta})]^2,$$

而由于估计量 $\overline{\lambda}(b)$ 和 $\lambda(b)$ 线性的, 以及由引理的条件, 有

$$\mathbf{E}[\alpha - \overline{\lambda}(\widetilde{\beta})]^2 = \mathbf{E}[\widetilde{\alpha} - \overline{\lambda}(\widetilde{\beta})]^2 \geqslant \mathbf{E}[\widetilde{\alpha} - \lambda(\widetilde{\beta})]^2 = \mathbf{E}[\alpha - \lambda(\beta)]^2,$$

于是, 证明了 $\lambda(\beta)$ 在线性估计类中的最优性. 最后得

$$\mathbf{E}\lambda(\beta) = \mathbf{E}\lambda(\widetilde{\beta}) = \mathbf{E}[\mathbf{E}(\widetilde{\alpha}|\widetilde{\beta})] = \mathbf{E}\widetilde{\alpha} = \mathbf{E}\alpha. \qquad \square$$

证明定理 2 与方程组 (13) 同时, 考虑方程组:

$$\begin{aligned}\widetilde{\theta}_{n+1} &= a_0 + a_1\widetilde{\theta}_n + a_2\widetilde{\xi}_n + b_1\widetilde{\varepsilon}_{11}(n+1) + b_2\widetilde{\varepsilon}_{12}(n+1), \\ \widetilde{\xi}_{n+1} &= A_0 + A_1\widetilde{\theta}_n + A_2\widetilde{\xi}_n + B_1\widetilde{\varepsilon}_{21}(n+1) + B_2\widetilde{\varepsilon}_{22}(n+1),\end{aligned} \tag{15}$$

其中 $\widetilde{\varepsilon}_{ij}(n)$ 是独立高斯随机变量, $\mathbf{E}\widetilde{\varepsilon}_{ij}(n) = 0$ 和 $\mathbf{E}\widetilde{\varepsilon}_{ij}^2(n) = 1$. 假设 $(\widetilde{\theta}_0, \widetilde{\xi}_0)$ 是高斯随机向量, 有与 (θ_0, ξ_0)相同的一阶矩和协方差, 而且不依赖于 $\widetilde{\varepsilon}_{ij}(n)$. 那么, 由于方程组 (15) 是线性的, 向量 $(\widetilde{\theta}_0, \cdots, \widetilde{\theta}_n, \widetilde{\xi}_0, \cdots, \widetilde{\xi}_n)$ 是高斯的, 即由引理 2 (确切一点说, 由定理 2 的明显的类似), 以及正态相关定理, 可以得到定理 2 的结论. $\square$

3. 例 下面是应用定理 1 和定理 2 的例子.

例 1 设 $\theta=(\theta_n)$ 和 $\eta=(\eta_n)$ 是两个 (弱) 平稳不相关随机序列, $\mathbf{E}\theta_n=\mathbf{E}\eta_n=0$, 而其谱密度相应为

$$f_\theta(\lambda)=\frac{1}{2\pi}\frac{1}{|1+b_1\mathrm{e}^{-i\lambda}|^2},\quad f_\eta(\lambda)=\frac{1}{2\pi}\frac{1}{|1+b_2\mathrm{e}^{-i\lambda}|^2},$$

其中 $|b_1|<1, |b_2|<1$.

下面将把 θ 视为有用信号, 而 η 视为噪声, 并且假设对序列 $\xi=(\xi_n)$ 进行观测:

$$\xi_n=\theta_n+\eta_n.$$

根据 §3 定理 3 的系 2, 存在 (互不相关的) 白噪声 $\varepsilon_1=(\varepsilon_1(n))$ 和 $\varepsilon_2=(\varepsilon_2(n))$, 使

$$\theta_{n+1}+b_1\theta_n=\varepsilon_1(n+1),\quad \eta_{n+1}+b_2\eta_n=\varepsilon_2(n+1).$$

那么,

$$\begin{aligned}\xi_{n+1}=\theta_{n+1}+\eta_{n+1}&=-b_1\theta_n-b_2\eta_n+\varepsilon_1(n+1)+\varepsilon_2(n+1)\\&=-b_2(\theta_n+\eta_n)-\theta_n(b_1-b_2)+\varepsilon_1(n+1)+\varepsilon_2(n+1)\\&=-b_2\xi_n-(b_1-b_2)\theta_n+\varepsilon_1(n+1)+\varepsilon_2(n+1).\end{aligned}$$

这样, 对于 θ 和 ξ 有递推公式

$$\begin{aligned}\theta_{n+1}&=-b_1\theta_n+\varepsilon_1(n+1),\\\xi_{n+1}&=-(b_1-b_2)\theta_n-b_2\xi_n+\varepsilon_1(n+1)+\varepsilon_2(n+1).\end{aligned}\tag{16}$$

而根据定理 2, $\widehat{m}_n=\widehat{\mathbf{E}}(\theta_n|\xi_0,\xi_1,\cdots,\xi_n)$ 和 $\widehat{\gamma}=\widehat{\mathbf{E}}(\theta_n-\widehat{m}_n)^2$ 满足如下最优线性滤波器的递推方程组:

$$\begin{aligned}m_{n+1}&=-b_1m_n+\frac{b_1(b_1-b_2)\gamma_n}{2+(b_1-b_2)^2\gamma_n}[\xi_{n+1}+(b_1-b_2)m_n+b_2\xi_n],\\\gamma_{n+1}&=b_1^2\gamma_n+1-\frac{[1+b_1(b_1-b_2)\gamma_n]^2}{2+(b_1-b_2)^2\gamma_n}.\end{aligned}\tag{17}$$

现在求为解该方程组所需要的初始条件 m_0 和 γ_0. 记 $d_{11}=\mathbf{E}\theta_n^2, d_{12}=\mathbf{E}\theta_n\xi_n$, $d_{22}=\mathbf{E}\xi_n^2$, 那么由 (16) 式, 可见

$$\begin{aligned}&d_{11}=b_1^2d_{11}+1,\\&d_{12}=b_1(b_1-b_2)d_{11}+b_1b_2d_{12}+1,\\&d_{22}=(b_1-b_2)^2d_{11}+b_2^2d_{22}+2b_2(b_1-b_2)d_{12}+2,\end{aligned}$$

因此,

$$d_{11}=\frac{1}{1-b_1^2},\quad d_{12}=\frac{1}{1-b_1^2},\quad d_{22}=\frac{2-b_1^2-b_2^2}{(1-b_1^2)(1-b_2^2)},$$

从而, 由 (14) 式得如下初始数据的值:

$$
\begin{aligned}
m_0 &= \frac{d_{12}}{d_{22}}\xi_0 = \frac{1-b_2^2}{2-b_1^2-b_2^2}\xi_0, \\
\gamma_0 &= d_{11} - \frac{d_{12}^2}{d_{22}} = \frac{1}{1-b_1^2} - \frac{1-b_2^2}{(1-b_1^2)(2-b_1^2-b_2^2)} = \frac{1}{2-b_1^2-b_2^2}.
\end{aligned} \tag{18}
$$

这样, 由 $\xi_0, \xi_1, \cdots, \xi_n$ 对信号 θ_n 的最优 (均方) 线性估计 m_n 和均方误差 γ_n, 决定于递推方程组 (17), 并且对于初始条件 (18) 求解. 需要指出, γ_n 的方程不含随机分量, 从而, 求 m_n 的数值所必须的 γ_n 的值可以事先 (在解该滤波之前) 计算出来.

例 2 这个例子从下面的角度可以借鉴: 它说明定理 2 的结果, 在序列 (θ, ξ) 服从不同于方程组 (13) 的 (非线性) 方程组时, 如何用于寻找最优线性滤波器的问题.

设 $\varepsilon_1 = (\varepsilon_1(n))$ 和 $\varepsilon_2 = (\varepsilon_2(n))$ 是两个独立高斯随机序列, 由独立随机变量构成, 且 $\mathbf{E}\varepsilon_i(n) = 0, \mathbf{E}\varepsilon_i^2(n) = 1, n \geqslant 1$. 考虑随机序列偶 $(\theta, \xi) = (\theta_n, \xi_n), n \geqslant 0$, 其中

$$
\begin{aligned}
\theta_{n+1} &= a\theta_n + (1+\theta_n)\varepsilon_1(n+1), \\
\xi_{n+1} &= A\theta_n + \varepsilon_2(n+1).
\end{aligned} \tag{19}
$$

假设 θ_0 不依赖于 $(\varepsilon_1, \varepsilon_2)$, 且 $\theta_0 \sim N(m_0, \gamma_0)$.

方程组 (19) 是非线性的, 故不能直接运用定理 2. 然而, 如果设

$$
\widetilde{\varepsilon}_1(n+1) = \frac{1+\theta_n}{\sqrt{\mathbf{E}(1+\theta_n)^2}}\varepsilon_1(n+1),
$$

则易见 $\mathbf{E}\widetilde{\varepsilon}_1(n) = 0, \mathbf{E}\widetilde{\varepsilon}_1(n)\widetilde{\varepsilon}_1(m) = 0 (m \neq n), \mathbf{E}\widetilde{\varepsilon}_1^2(n) = 1$. 所以与 (19) 式同时, 原序列 (θ, ξ) 也满足线性方程组

$$
\begin{aligned}
\theta_{n+1} &= a_1\theta_n + b_1(n)\widetilde{\varepsilon}_1(n+1), \\
\xi_{n+1} &= A\theta_n + \varepsilon_2(n+1),
\end{aligned} \tag{20}
$$

其中 $b_1(n) = \sqrt{\mathbf{E}(1+\theta_n)^2}$, 而 $\{\widetilde{\varepsilon}_1(n)\}$ 是某一两两不相关随机变量序列.

方程组 (20) 是形如 (13) 的线性方程组, 即根据定理 2, 最优线性估计量 $\widehat{m}_n = \widehat{\mathbf{E}}(\theta_n | \xi_0, \xi_1, \cdots, \xi_n)$ 及其误差 $\widehat{\gamma}$ 可以由方程组 (7), (8) 决定, 在这种情形下方程组 (7), (8) 有如下形式:

$$
\begin{aligned}
\widehat{m}_{n+1} &= a_1\widehat{m}_n + \frac{a_1 A_1 \widehat{\gamma}_n}{1 + A_1^2\widehat{\gamma}_n}[\xi_{n+1} - A_1\widehat{m}_n], \\
\widehat{\gamma}_{n+1} &= [a_1^2\widehat{\gamma}_n + b_1^2(n)] - \frac{(a_1 A_1 \widehat{\gamma}_n)^2}{1 + A_1^2\widehat{\gamma}_n},
\end{aligned}
$$

其中 $b_1(n) = \sqrt{\mathbf{E}(1+\theta_n)^2}$ 应由方程组 (19) 的第一个方程来求.

例 3 **参数估计** 设 $\theta = (\theta_1, \cdots, \theta_k)$ 是高斯向量, 且 $\mathbf{E}\theta = m$ 和 $\operatorname{cov}(\theta, \theta) = \gamma$. 假设 $\xi = (\xi_n), n \geqslant 0$, 是 l 维随机序列, 且

$$\xi_{n+1} = A_0(n, \xi) + A_1(n, \xi)\theta + B_1(n, \xi)\varepsilon_1(n+1), \quad \xi_0 = 0, \tag{21}$$

其中 ε_1 的含义与方程组 (1) 中的含义相同. 现在对于已知的 m 和 γ, 欲根据对 ξ 的观测结果, 求 θ 的最优估计.

那么, 对于 $m_n = \mathbf{E}(\theta|\mathscr{F}_n^\xi)$ 和 γ_n, 由 (7), (8) 式, 有

$$\begin{aligned} m_{n+1} = m_n + \gamma_n A_1^*(n, \xi)[(B_1 B_1^*)(n, \xi) + A_1(n, \xi)\gamma_n A_1^*(n, \xi)]^{\oplus} \\ \times[\xi_{n+1} - A_0(n, \xi) - A_1(n, \xi)m_n], \\ \gamma_{n+1} = \gamma_n - \gamma_n A_1^*(n, \xi)[(B_1 B_1^*)(n, \xi) + A_1(n, \xi)\gamma_n A_1^*(n, \xi)]^{\oplus} A_1(n, \xi)\gamma_n. \end{aligned} \tag{22}$$

如果 $B_1 B_1^*$ 是非退化矩阵, 则方程组 (22) 的解为

$$\begin{aligned} m_{n+1} &= \left[E + \gamma \sum_{i=0}^{n} A_1^*(i, \xi)(B_1 B_1^*)^{-1}(i, \xi) A_1^*(i, \xi)\right]^{-1} \times \\ &\quad \left[m + \gamma \sum_{i=0}^{n} A_1^*(i, \xi)(B_1 B_1^*)^{-1}(i, \xi)(\xi_{i+1} - A_0(m, \xi))\right], \\ \gamma_{n+1} &= \left[E + \gamma \sum_{i=0}^{n} A_1^*(i, \xi)(B_1 B_1^*)^{-1}(i, \xi) A_1(i, \xi)\right]^{-1} \gamma, \end{aligned} \tag{23}$$

其中 E 是单位矩阵.

4. 练习题

1. 证明, 对于概型 (1), 向量 m_n 和 $\theta_n - m_n$ 不相关:

$$\mathbf{E} m_n^*(\theta_n - m_n) = 0.$$

2. 设概型 (1) 中的 γ_0 和所有系数 (或许除系数 $a_0(n, \xi), A_0(n, \xi)$ 外) 都与 "随机性" 无关 (即不依赖于 ξ). 证明条件协方差 γ_n 也与 "随机性" 无关: $\gamma_n = \mathbf{E}\gamma_n$.

3. 证明方程组 (22) 的解由公式 (23) 表示.

4. 设 $(\theta, \xi) = (\theta_n, \xi_n)$ 是高斯序列, 满足概型 (1) 的如下特别形式:

$$\theta_{n+1} = a\theta_n + b\varepsilon_1(n+1), \quad \xi_{n+1} = A\theta_n + B\varepsilon_2(n+1).$$

证明, 如果 $A \neq 0, b \neq 0, B \neq 0$, 则极限过滤误差 $\gamma = \lim\limits_{n\to\infty} \gamma_n$ 存在, 且为方程

$$\gamma^2 + \left[\frac{B^2(1-a^2)}{A^2} - b^2\right]\gamma - \frac{b^2 B^2}{A^2} = 0$$

的正根.

5. (内插; [41,13.3]). 设 (θ,ξ) 是部分可观测序列, 且服从递推关系式 (1) 和 (2). 设向量 θ_m 的条件分布

$$\pi_a(m,m) = \mathbf{P}\{\theta_m \leqslant a | F_m^{\xi}\}$$

是正态的.

(a) 证明, 对于 $n \geqslant m$, 条件分布

$$\pi_a(m,n) = \mathbf{P}\{\theta_m \leqslant a | \mathscr{F}_n^{\xi}\}$$

也是正态的, $\pi_a(m,n) \sim N(\mu(m,n),\gamma(m,n))$.

(b) 求随机变量 θ_m (关于 $\mathscr{F}_n^{\xi}$) 的内插估计 $\mu(m,n)$, 和矩阵 $\gamma(m,n)$.

6. (外推; [41,13.4]). 设关系式 (1) 和 (2) 中的变量为:

$$a_0(n,\xi) = a_0(n) + a_2(n)\xi_n, \quad a_1(n,\xi) = a_1(n),$$
$$A_0(n,\xi) = A_0(n) + A_2(n)\xi_n, \quad A_1(n,\xi) = A_1(n).$$

(a) 证明, 在这种情形下 $(n \geqslant m)$, 分布 $\pi_{a,b}(m,n) = \mathbf{P}\{\theta_m \leqslant a, \xi_n \leqslant b | \mathscr{F}_n^{\xi}\}$ 是正态的.

(b) 求外推估计量

$$\mathbf{E}(\theta_m | \mathscr{F}_m^{\xi}) \quad 和 \quad \mathbf{E}(\xi_n | \mathscr{F}_m^{\xi}).$$

7. (最优控制; [41,14.3]). 考虑 "被控制" 部分观测系统 $(\theta_n,\xi_n)_{0\leqslant n\leqslant N}$, 其中

$$\theta_{n+1} = u_n + \theta_n + b\varepsilon_1(n+1),$$
$$\xi_{n+1} = \theta_n + \varepsilon_2(n+1).$$

这里, "控制" u_n 为 $\mathscr{F}_n^{\xi}$– 可测, 并且对于一切 $0 \leqslant n \leqslant N-1, \mathbf{E}u_n^2 < \infty$. 随机变量 $\varepsilon_1(n)$ 和 $\varepsilon_2(n), n = 1,\cdots,N$ 的含义与方程组 (1) 及等式 (2) 相同; $\xi_0 = 0, \theta_0 \sim N(m,\gamma)$.

我们称 "控制" $u^* = (u_0^*,\cdots,u_{N-1}^*)$ 是最优的, 如果

$$V(u^*) = \sup_u V(u), \quad 其中 \quad V(u) = \mathbf{E}\left[\sum_{n=0}^{N-1}(\theta_n^2 + u_n^2) + \theta_N^2\right].$$

证明:

$$u_n^* = -[1+P_{n+1}]^+ P_{n+1} m_n^*, \quad n = 0,\cdots,N-1,$$

其中

$$a^+ = \begin{cases} a^{-1}, & a \neq 0, \\ 0, & a = 0; \end{cases}$$

且 $(P_n)_{0\leqslant n\leqslant N}$ 决定于如下递推关系式:

$$P_n = 1 + P_{n+1} - P_{n+1}^2[1 + P_{n+1}]^+, \quad P_N = 1,$$

而 (m_n^*) 决定于关系式

$$m_{n+1}^* = u_n^* + \gamma_n^*(1 + \gamma_n^*)^+(\xi_{n+1} - m_n^*), \quad 0 \leqslant n \leqslant N - 1,$$

其中 $m_0^* = m$, 而 $\gamma_0^* = \gamma$,

$$\gamma_{n+1}^* = \gamma_n^* + 1 - (\gamma_n^*)^2(1 + \gamma_n^*)^+, \quad 0 \leqslant n \leqslant N - 1.$$



第七章　构成鞅的随机变量序列

[1] 按 "全国科学技术名词审定委员会" 审定公布定名: submartingale 为 "下鞅". —— 译者



鞅论很好地描绘数学概率论形成的历史 —— 鞅论的基本概念, 是受博弈实践的启发引进的, 但是后来这些概念成为现代数学非常精细的工具之一 ······

J. L. 杜布 (J. L. Doob)《什么是鞅?》[127]

---

§1. 鞅和相关概念的定义

1. 这一节的研究对象 —— 鞅 在概率论中, 用各种不同的方法研究随机变量之间的相依性. 在 (弱) 随机序列的理论中, 基本特征是*协方差函数*, 而且该理论的一切结论完全决定于协方差函数的性质. 在马尔可夫链的理论中 (第一章 §12 和第八章), 基本特征是*转移函数*, 它决定由马尔可夫相依性联系的随机变量的演变.

在这一章 (亦见第一章 §11) 将引进相当广泛的随机变量序列类 (鞅及其推广), 而关于其相依性的研究, 利用基于*条件数学期望*性质的研究方法.

2. 鞅 (下鞅) 的概念 假设给定一带过滤 (流) 的概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$, 所谓过滤 (流) 即 σ - 代数族 $(\mathscr{F}_n)$, 其中 $\mathscr{F}_n, n \geqslant 0$, 而且 $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \cdots \subseteq \mathscr{F}$ ("过滤概率空间").

设 $X_0, X_1, \cdots$ 是概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 上的随机变量. 如果对于每一个 $n \geqslant 0$, 变量 X_n 是 $\mathscr{F}_n$ - *可测的*, 则组合 $X = (X_n, \mathscr{F}_n)_{n\geqslant 0}$ 或简记 $X = (X_n, \mathscr{F}_n)$, 称之为*随机序列*.

假如随机序列 $X = (X_n, \mathscr{F}_n)$ 还具有性质: 每一个 $n \geqslant 1$, 变量 X_n 是 $\mathscr{F}_{n-1}$ - *可测的*, 则记作 $X = (X_n, \mathscr{F}_{n-1})$, 并且设 $\mathscr{F}_{-1} = \mathscr{F}_0$. 这时称 X 为*可预测随机序列*. 序列 $(X_n)_{n\geqslant 0}$ 称做递增的, 如果 $X_0 = 0$ 且 $X_n \leqslant X_{n+1} (\mathbf{P}$ - a.c.).

定义 1 随机序列 $X = (X_n, \mathscr{F}_n)$ 称做*鞅* (*下鞅*), 如果对于一切 $n \geqslant 0$, 有,

$$\mathbf{E}|X_n| < \infty, \tag{1}$$

$$\mathbf{E}(X_{n+1}|\mathscr{F}_n) \underset{(\geqslant)}{=} X_n \quad (\mathbf{P} - \text{a.c.}). \tag{2}$$

随机序列 $X = (X_n, \mathscr{F}_n)$ 称做*上鞅*, 如果序列 $-X = (-X_n, \mathscr{F}_n)$ 是*下鞅*.

在特殊情形下, 假如 $\mathscr{F}_n = \mathscr{F}_n^X$, 其中 $\mathscr{F}_n^X = \sigma(X_0, \cdots, X_n)$, 并且随机序列 $X = (X_n, \mathscr{F}_n)$ 是鞅 (下鞅), 则称序列 $(X_n)_{n\geqslant 0}$ 本身是鞅 (下鞅).

由条件数学期望的性质, 容易证明, 条件 (2) 等价于: 对于任何 $n \geqslant 0$ 和 $A \in \mathscr{F}_n$, 有

$$\int_A X_{n+1} \mathrm{d}\mathbf{P} \underset{(\geqslant)}{=} \int_A X_n \mathrm{d}\mathbf{P}. \tag{3}$$

例 1 设 $(\xi_n)_{n\geqslant 0}$ 是独立随机变量序列, $\mathbf{E}|\xi_n| < \infty, \mathbf{E}\xi_n = 0, X_n = \xi_0 + \cdots + \xi_n, \mathscr{F}_n = \sigma(\xi_0, \cdots, \xi_n)$, 则随机序列 $X = (X_n, \mathscr{F}_n)$ 是鞅.

例 2 设 $(\xi_n)_{n\geqslant 0}$ 是独立随机变量序列, $\mathbf{E}\xi_n = 1$, 则随机序列 $X = (X_n, \mathscr{F}_n)$ 也是鞅, 其中

$$X_n = \prod_{k=0}^{n} \xi_k, \quad \mathscr{F}_n = \sigma(\xi_0, \cdots, \xi_n).$$

例 3 设 ξ 是随机变量序列, $\mathbf{E}|\xi| < \infty$, 而 $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \cdots \subseteq \mathscr{F}$. 则随机序列 $X = (X_n, \mathscr{F}_n)$ 是鞅, 其中 $X_n = (\xi|\mathscr{F}_n)$ (称做列维 (P. P. Lévy) 鞅).

例 4 如果 $(\xi_n)_{n\geqslant 0}$ 是非负可积随机变量序列, 则序列 (X_n) 是鞅, 其中 $X_n = \xi_0 + \cdots + \xi_n$.

例 5 设 $X = (X_n, \mathscr{F}_n)$ 是鞅, 而 $g(x)$ 是凹 (向下凸) 函数, 且 $\mathbf{E}|g(X_n)| < \infty, n \geqslant 0$, 则 (由第二章 §6 的延森 (I. L. Iensen) 不等式可见) 随机序列 $(g(X_n), \mathscr{F}_n)$ 是下鞅.

假如 $X = (X_n, \mathscr{F}_n)$ 是下鞅, 而 $g(x)$ 是凹函数, 且对于一切, $n \geqslant 0, \mathbf{E}|g(X_n)| < \infty$, 则随机序列 $(g(X_n), \mathscr{F}_n)$ 也是下鞅.

在定义 1 中所作的假设 $\mathbf{E}|X_n| < \infty$, 保障条件数学期望 $\mathbf{E}(X_{n+1}|\mathscr{F}_n), n \geqslant 0$, 的存在性. 不过, 在不要求 $\mathbf{E}|X_{n+1}| < \infty$ 的情况下, 这些条件数学期望也可能存在. 我们知道根据第二章 §7, $\mathbf{E}(X_{n+1}^+|\mathscr{F}_n)$ 和 $\mathbf{E}(X_{n+1}^-|\mathscr{F}_n)$ 总有定义, 如果 (若 $\mathbf{P}(A\triangle B) = 0$, 则以 $A = B$ ($\mathbf{P}$ – a.c.) 表示)

$$\{\omega : \mathbf{E}(X_{n+1}^+|\mathscr{F}_n) < \infty\} \bigcup \{\omega : \mathbf{E}(X_{n+1}^-|\mathscr{F}_n) < \infty\} = \Omega \quad (\mathbf{P} - \text{a.c.}),$$

那么称 $\mathbf{E}(X_{n+1}|\mathscr{F}_n)$ 有定义, 且根据定义设

$$\mathbf{E}(X_{n+1}|\mathscr{F}_n) = \mathbf{E}(X_{n+1}^+|\mathscr{F}_n) - \mathbf{E}(X_{n+1}^-|\mathscr{F}_n).$$

由此可见, 下面的定义是自然的.

定义 2 随机序列 $X = (X_n, \mathscr{F}_n)$ 称做广义鞅 (下鞅), 如果对于一切 $n \geqslant 0$, 有 $\mathbf{E}|X_0| < \infty$, 而且 $\mathbf{E}(X_{n+1}|\mathscr{F}_n)$ 有定义, 并满足条件 (2).

注意, 由该定义可见, 对于广义下鞅 $\mathbf{E}(X_{n+1}^-|\mathscr{F}_n) < \infty$, 而对于广义鞅 $\mathbf{E}(|X_{n+1}||\mathscr{F}_n) < \infty$ ($\mathbf{P}$–a.c.).

3. 马尔可夫时间 下面的定义引进的马尔可夫时间的概念, 在以后介绍的全部理论中有十分重要的作用.

定义 3 在集合 $\{0, 1, \cdots, +\infty\}$ 中取值的随机变量 $\tau = \tau(\omega)$, 称做 (关于 σ – 代数族 $(\mathscr{F}_n)$ 的) 马尔可夫时间或"不依赖于将来的随机变量", 如果对于每一个 $n \geqslant 0$,

$$\{\tau = n\} \in \mathscr{F}_n. \tag{4}$$

对于 $\mathbf{P}\{\tau < \infty\} = 1$ 的情形, 马尔可夫时间 τ, 称做停止时间.

设 $X=(X_n,\mathscr{F}_n)$ 是一随机序列, 而 τ (关于 σ – 代数族 $(\mathscr{F}_n)$) 是马尔可夫时间. 记

$$X_\tau(\omega)=\sum_{n=0}^{\infty}X_n(\omega)I_{\{\tau=n\}}(\omega)$$

(于是 $X_\infty=0$, 即在集合 $\{\omega:\tau=\infty\}$ 上 $X_\tau=0$).

那么, 对于每一个 $B\in\mathscr{B}(\mathbb{R})$,

$$\{\omega:X_\tau\in B\}=\{\omega:X_\infty\in B,\tau=\infty\}+\sum_{n=0}^{\infty}\{\omega:X_n\in B,\tau=n\}\in\mathscr{F},$$

从而 $X_\tau=X_{\tau(\omega)}(\omega)$ 是随机变量.

例 6 设 $X=(X_n,\mathscr{F}_n)$ 是一随机序列, 而 $B\in\mathscr{B}(\mathbb{R})$. 那么, (**首次到达集合** B **中的**) 时间

$$\tau_B=\inf\{n\geqslant 0:X_n\in B\}$$

(当 $\{\cdot\}=\varnothing$ 时 $\tau_B=+\infty$) 是马尔可夫时间, 因为对于任何 $n\geqslant 0$, 有

$$\{\tau_B=n\}=\{X_0\notin B,\cdots,X_{n-1}\notin B,X_n\in B\}\in\mathscr{F}_n.$$

例 7 设 $X=(X_n,\mathscr{F}_n)$ 是鞅 (下鞅), 而 τ (关于 σ – 代数族 $(\mathscr{F}_n)$) 是马尔可夫时间. 那么, "停止" 序列 $X^\tau=(X_{n\wedge\tau},\mathscr{F}_n)$ 也构成鞅 (下鞅).

事实上, 由关系式

$$X_{n\wedge\tau}=\sum_{m=0}^{n-1}X_mI_{\{\tau=m\}}=X_nI_{\{\tau\geqslant n\}}$$

可见, 随机变量 $X_{n\wedge\tau}$ 为 $\mathscr{F}_n$ – 可测的和可积的, 且

$$X_{(n+1)\wedge\tau}-X_{n\wedge\tau}=I_{\{\tau>n\}}(X_{n+1}-X_n),$$

由此可见

$$\mathbf{E}[X_{(n+1)\wedge\tau}-X_{n\wedge\tau}|\mathscr{F}_n]=I_{\{\tau>n\}}\mathbf{E}[X_{n+1}-X_n|\mathscr{F}_n]\underset{(\geqslant)}{=}0.$$

与每一个 σ – 代数族 $(\mathscr{F}_n)$, 及关于 $(\mathscr{F}_n)$ 的马尔可夫时间 τ, 可以与一集系

$$\mathscr{F}_\tau=\{A\in\mathscr{F}:A\bigcap\{\tau=n\}\in\mathscr{F}_n,\ \text{对于一切}\ n\geqslant 0\}$$

相联系. 显然 $\Omega\in\mathscr{F}_\tau$ 和 $\mathscr{F}_\tau$ 关于求可数次并运算封闭. 此外, 如果 $A\in\mathscr{F}_\tau$, 则

$$\overline{A}\bigcap\{\tau=n\}=\{\tau=n\}\backslash(A\bigcap\{\tau=n\})\in\mathscr{F}_n,$$

故 $\overline{A}\in\mathscr{F}_\tau$. 由此可见 $\mathscr{F}_\tau$是 σ – 代数.

如果把 $\mathscr{F}_n$ 视为在时刻 n (包括 n) 之前观测到的事件的全体, 那么 $\mathscr{F}_\tau$ 可以视为在 "随机" 时间 τ 观测到的事件的全体.

不难证明 (练习题 3), 随机变量 τ 和 X_τ 是 $\mathscr{F}_\tau$ – 可测的.

4. 局部鞅和鞅变换

定义 4 称随机序列 $X = (X_n, \mathscr{F}_n)$ 为局部鞅 (局部下鞅), 如果存在 (局部化的) 有限马尔可夫时间序列 $(\tau_k)_{k\geqslant 1} : \tau_k \leqslant \tau_{k+1}(\mathbf{P} - \text{a.c.})$, $\tau_k \uparrow \infty(\mathbf{P} - \text{a.c.})$, $k \to \infty$, 且每一个 ("停止") 序列 $X^{\tau_k} = (X_{\tau_k \wedge n} I_{\{\tau_k > 0\}}, \mathscr{F}_n)$ 是鞅 (下鞅).

下面在定理 1 中证明, 实际上局部鞅类与广义鞅类等同. 此外, 每一局部鞅可以借助于所谓鞅变换, 由某一鞅和某一可预测序列得到.

定义 5 设 $Y = (Y_n, \mathscr{F}_n)_{n\geqslant 0}$ 是随机序列, 而 $V = (V_n, \mathscr{F}_n)_{n\geqslant 0}$ 是可预测序列 $(\mathscr{F}_{-1} = \mathscr{F}_0)$. 随机序列 $V \cdot Y = ((V \cdot Y)_n, \mathscr{F}_n)$, 其中

$$(V \cdot Y)_n = V_0 Y_0 + \sum_{i=1}^{n} V_i \Delta Y_i, \tag{5}$$

而 $\Delta Y_i = Y_i - Y_{i-1}$, 称做 Y 由 V 的变换. 此外, 如果 Y 是鞅 (或局部鞅), 则称 $V \cdot Y$ 为鞅变换.

定理 1 设 $X = (X_n, \mathscr{F}_n)_{n\geqslant 0}$ 是随机序列, 其中 $X_0 = 0(\mathbf{P} - \text{a.c.})$. 下列各条件等价:

a) X 是局部鞅;

b) X 是广义鞅;

c) X 是鞅变换, 即存在可预测序列 $V = (V_n, \mathscr{F}_n)_{n\geqslant 0}(V_0 = 0)$, 和鞅 $Y = (Y_n, \mathscr{F}_n)_{n\geqslant 0}(Y_0 = 0)$, 使 $X = V \cdot Y$.

证明 a) $\Rightarrow$ b). 设 X 是局部鞅, 而 $(\tau_k)_{k\geqslant 1}$ 是局部化马尔可夫时间序列. 那么, 对于任何 $m \geqslant 0$,

$$\mathbf{E}[|X_{m\wedge\tau_k}| I_{\{\tau_k > 0\}}] < \infty, \tag{6}$$

从而

$$\mathbf{E}[|X_{(n+1)\wedge\tau_k}| I_{\{\tau_k > n\}}] = \mathbf{E}[|X_{n+1}| I_{\{\tau_k > n\}}] < \infty. \tag{7}$$

随机变量 $I_{\{\tau_k > n\}}$ 是 $\mathscr{F}_n-$ 可测的. 因此由 (7) 式可见, 有

$$\mathbf{E}[|X_{n+1}| I_{\{\tau_k > n\}} | \mathscr{F}_n] = I_{\{\tau_k > n\}} \mathbf{E}[|X_{n+1}| | \mathscr{F}_n] < \infty\ (\mathbf{P} - \text{a.c.}).$$

这里, $I_{\{\tau_k > n\}} \to 0\ (\mathbf{P} - \text{a.c.})$, $k \to \infty$, 因而

$$\mathbf{E}[|X_{n+1}| | \mathscr{F}_n] < \infty \quad (\mathbf{P} - \text{a.c.}). \tag{8}$$

由于这一条件, 知 $\mathbf{E}[X_{n+1} | \mathscr{F}_n]$ 有定义, 故只剩下证明 $\mathbf{E}[X_{n+1} | \mathscr{F}_n] = X_n\ (\mathbf{P} - \text{a.c.})$.

为此, 需要对于 $A \in \mathscr{F}_n$, 证明

$$\int_A X_{n+1} \mathrm{d}\mathbf{P} = \int_A X_n \mathrm{d}\mathbf{P}.$$

由第二章 §7 练习题 7 知, $\mathbf{E}[|X_{n+1}||\mathscr{F}_n] < \infty(\mathbf{P}-\text{a.c.})$, 当且仅当测度

$$\int_A X_{n+1}\mathrm{d}\mathbf{P}, \quad A \in \mathscr{F}_n$$

是 $\sigma-$ 有限的. 现在证明测度 $\int_A |X_n|\mathrm{d}\mathbf{P}, A \in \mathscr{F}_n$, 也是 $\sigma-$ 有限的.

由于 X^{τ_k} 是鞅, 故 $|X^{\tau_k}| = (|X_{\tau_k \wedge n}|I_{\{\tau_k>0\}}, \mathscr{F}_n)$ 是下鞅, 从而 $(\{\tau_k > n\} \in \mathscr{F}_n)$,

$$\begin{aligned}\int_{A\cap\{\tau_k>n\}} |X_n|\mathrm{d}\mathbf{P} &= \int_{A\cap\{\tau_k>n\}} |X_{n\wedge\tau_k}|I_{\{\tau_k>0\}}\mathrm{d}\mathbf{P}\\ &\leqslant \int_{A\cap\{\tau_k>n\}} |X_{(n+1)\wedge\tau_k}|I_{\{\tau_k>0\}}\mathrm{d}\mathbf{P} = \int_{A\cap\{\tau_k>n\}} |X_{n+1}|\mathrm{d}\mathbf{P}.\end{aligned}$$

令 $k\to\infty$, 得

$$\int_A |X_n|\mathrm{d}\mathbf{P} \leqslant \int_A |X_{n+1}|\mathrm{d}\mathbf{P},$$

因此就证明了测度 $\int_A X_{n+1}\mathrm{d}\mathbf{P}, A \in \mathscr{F}$, 是 $\sigma-$ 有限的.

设 $A \in \mathscr{F}$, 使 $\int_A |X_{n+1}|\mathrm{d}\mathbf{P} < \infty$. 那么, 根据勒贝格控制收敛定理, 可以在下列式子中求极限:

$$\int_{A\cap\{\tau_k>n\}} X_n\mathrm{d}\mathbf{P} = \int_{A\cap\{\tau_k>n\}} X_{n+1}\mathrm{d}\mathbf{P},$$

而由于 X 是局部鞅, 可见求极限是合理的. 从而, 有

$$\int_A X_n\mathrm{d}\mathbf{P} = \int_A X_{n+1}\mathrm{d}\mathbf{P},$$

而且对于任意 $A \in \mathscr{F}_n, \int_A |X_{n+1}|\mathrm{d}\mathbf{P} < \infty$. 由此可见, 对于任意 $A \in \mathscr{F}_n$, 上面的等式成立, 从而 $\mathbf{E}[X_{n+1}|\mathscr{F}_n] = X_n(\mathbf{P}-\text{a.c.})$.

b) $\Rightarrow$ c). 设 $\Delta X_n = X_n - X_{n-1}, X_0 = 0; V_0 = 0, V_n = \mathbf{E}[|\Delta X_n||\mathscr{F}_{n-1}](n \geqslant 1)$; 记

$$W_n = V_n^{\oplus}\left(= \begin{cases} V_n^{-1}, & V_n \neq 0,\\ 0, & V_n \neq 0,\end{cases}\right),$$

$Y_0 = 0$ 和 $Y_n = \sum\limits_{i=1}^n W_i\Delta X_i(n \geqslant 1)$. 显然

$$\mathbf{E}[|\Delta Y_n||\mathscr{F}_{n-1}] \leqslant 1, \quad \mathbf{E}[\Delta Y_n|\mathscr{F}_{n-1}] = 0,$$

从而, $Y = (Y_n, \mathscr{F}_n)$ 是鞅. 其次 $X_0 = V_0 \cdot Y_0 = 0, \Delta(V\cdot Y)_n = \Delta X_n$. 于是 $X = V\cdot Y$.

c) $\Rightarrow$ a). 设 $X = V\cdot Y$, 其中 V 是可预测序列, Y 是鞅, 而 $V_0 = Y_0 = 0$. 记

$$\tau_k = \inf\{n \geqslant 0 : |V_{n+1}| > k\},$$

且当集合 $\{\cdot\}=\varnothing$ 时 $\tau_\infty=\infty$. 由于 V_{n+1} 是 $\mathscr{F}_n-$ 可测的, 可见对于每一个 $k\geqslant 1$, 变量 τ_k 是马尔可夫时间.

考虑序列 $X^{\tau_k}=((V\cdot Y)_{n\wedge\tau_k}I_{\{\tau_k>0\}},\mathscr{F}_n)$.

在集合 $\{\tau_k>0\}$ 上 $|V_{n\wedge\tau_k}|\leqslant k$. 故对于任意 $n\geqslant 1$, 有 $\mathbf{E}|(V\cdot Y)_{n\wedge\tau_k}I_{\{\tau_k>0\}}|<\infty$. 其次, 对于任意 $n\geqslant 1$, 有

$$\begin{aligned}&\mathbf{E}\{[(V\cdot Y)_{(n+1)\wedge\tau_k}-(V\cdot Y)_{n\wedge\tau_k}]I_{\{\tau_k>0\}}|\mathscr{F}_n\}\\=&I_{\{\tau_k>0\}}V_{(n+1)\wedge\tau_k}\mathbf{E}\{Y_{(n+1)\wedge\tau_k}-Y_{n\wedge\tau_k}|\mathscr{F}_n\}=0,\end{aligned}$$

因为 (见例 7) $\mathbf{E}\{Y_{(n+1)\wedge\tau_k}-Y_{n\wedge\tau_k}|\mathscr{F}_n\}=0$.

于是, 对于每一个 $k\geqslant 1$, "停止" 序列 X^{τ_k} 是鞅, 其中 $\tau_n\uparrow\infty$ $(\mathbf{P}-\text{a.c.})$, 从而 X 是局部鞅. □

5. 博弈与数学概念 "鞅" 的产生

例 8 $(\eta_n)_{n\geqslant 1}$ 是独立同分布的伯努利随机变量, 且 $\mathbf{P}\{\eta_n=1\}=p,\mathbf{P}\{\eta_n=-1\}=q,p+q=1$. 假如考虑某项博弈, 可以把事件 $\{\eta_n=1\}$ 视为某选手在第 n 局中 "成功" (赢), 而把事件 $\{\eta_n=-1\}$ 视为在第 n 局中 "失败" (输). 假设 V_n 是第 n 局的赌注. 那么, 赌徒经 n 局赢得的总金额 (总收益) 等于

$$X_n=\sum_{i=1}^n V_i\eta_i=X_{n-1}+V_n\eta_n,\quad X_0=0.$$

完全自然, 第 n 局的赌注 V_n 可依赖于上一局的结果, 即依赖于 $V_1,\cdots,V_{n-1}$ 和 $\eta_1,\cdots,\eta_{n-1}$. 换句话说, 如果设 $\mathscr{F}_0=(\varnothing,\Omega)$ 和 $\mathscr{F}_n=\sigma\{\eta_1,\cdots,\eta_n\}$, 则 V_n 是 $\mathscr{F}_{n-1}-$ 可测随机变量, 即决定赌徒 "策略" 的序列 $V=(V_n,\mathscr{F}_{n-1})$ 是可以预测的. 若设 $Y_n=\eta_1+\cdots+\eta_n$, 则有

$$X_n=\sum_{i=1}^n V_i\Delta Y_i,$$

则序列 $X=(X_n,\mathscr{F}_n)$, 其中 $X_0=0$, 是 Y 由 V 的变换.

按博弈的观点, 如果每一步的期望收益 $\mathbf{E}(X_{n+1}-X_n|\mathscr{F}_n)=0$ ($\geqslant 0$或 $\leqslant 0$), 则该博弈是公平的 (有利的或不利的). 因此显然, 博弈是

公平的, 如果 $p=q=1/2$,

有利的, 如果 $p>q$,

不利的, 如果 $p<q$.

由于序列 $X=(X_n,\mathscr{F}_n)$ 构成

鞅, 如果 $p=q=1/2$,

下鞅, 如果 $p>q$,

上鞅, 如果 $p < q$,

则可以认为, 关于该博弈是公平的 (有利的或不利的), 对应于关于序列 X 是鞅 (下鞅或上鞅).

现在考虑特别的 "策略" 类 $V_n = (V_n, \mathscr{F}_{n-1})_{n\geqslant 1}$, 其中 $V_1 = 1$ 和

$$V_n = \begin{cases} 2^{n-1}, & 若\ \eta_1 = -1, \cdots, \eta_{n-1} = -1, \\ 0, & 若不然, \end{cases} \quad n > 1, \tag{9}$$

该式的含义是: 赌徒从赌注 $V_1 = 1$ 开始, 每当他赢一局时就将赌注增加一倍, 而每当他输一局时就终止博弈.

假如 $\eta_1 = -1, \cdots, \eta_n = -1$, 则经 n 局赌徒的总损失等于

$$\sum_{i=1}^{n} 2^{i-1} = 2^n - 1.$$

因此, 假如再假设 $\eta_{n+1} = 1$, 则

$$X_{n+1} = X_n + V_{n+1} = -(2^n - 1) + 2^n = 1.$$

记 $\tau = \inf\{n \geqslant 1 : X_n = 1\}$. 如果 $p = q = 1/2$, 即所考虑的博弈是公平的, 则

$$\mathbf{P}\{\tau = n\} = \frac{1}{2^n}, \quad \mathbf{P}\{\tau < \infty\} = 1, \quad \mathbf{P}\{X_\tau = 1\} = 1, \quad \mathbf{E}X_\tau = 1.$$

这样, 甚至在公平博弈的情形下, 如果赌徒坚持运用 "策略" (9), 他 (以概率 1) 经过有限时间完全可以顺利的结束博弈, 并使自己的赌注再增加一个单位 ($\mathbf{E}X_\tau = 1 > X_0 = 0$).

在博弈的实际中, 在输局时加倍赌注而在首次赌赢时停止博弈, 所描绘博弈的系统称做鞅. 也正式由此产生了数学概念 "鞅".

注 在 $p = q = 1/2$ 的情形下, 序列 $X = (X_n, \mathscr{F}_n)_{n\geqslant 0}$ 是鞅, 其中 $X_0 = 0$, 因而对于任何 $n \geqslant 1$,

$$\mathbf{E}X_n = \mathbf{E}X_0 = 0.$$

因此, 假如将考虑时刻 n 换成考虑随机时刻 τ, 则可以指望该关系式仍然成立. 以后 (§2 定理 1) 会清楚地看到, 在 "典型" 的场合 $\mathbf{E}X_\tau = \mathbf{E}X_0$ 成立. 而这一等式不成立 (如上面考虑的博弈) 的情形, 出现在所谓 "物理上不可实现" 的场合, 这时或者 τ, 或者 $|X_n|$ 取的值过分地大. (应该指出, 上面考虑的博弈的情形是物理上无法实现的, 因为它假定博弈时间无限, 以及赌徒的初始赌资无限.)

6. 鞅 – 差

定义 6 随机序列 $\xi = (\xi_n, \mathscr{F}_n)_{n\geqslant 0}$ 称做鞅 – 差, 如果对于一切 $n \geqslant 0, \mathbf{E}|\xi_n| < \infty$ 且

$$\mathbf{E}(\xi_{n+1}|\mathscr{F}_n) = 0 \ (\mathbf{P} - \text{a.c.}). \tag{10}$$

由定义 1 和 6 可以清楚地看到鞅与鞅 - 差之间的联系. 具体地说, 如果 $X = (X_n, \mathscr{F}_n)$ 是鞅, 则 $\xi = (\xi_n, \mathscr{F}_n)$ 是鞅 - 差, 其中 $\xi_0 = X_0, \xi_n = \Delta X_n, n \geqslant 1$. 同样地, 如果 $\xi = (\xi_n, \mathscr{F}_n)$ 是鞅 - 差, 则 $X = (X_n, \mathscr{F}_n)$ 是鞅, 其中 $X_n = \xi_0 + \cdots + \xi_n$.

按照这一术语, 如果 $\mathbf{E}\xi_n = 0, n \geqslant 0$, 而相应的 $\sigma-$ 代数为 $\mathscr{F}_n = \sigma(\xi_0, \cdots, \xi_n)$, 则任何独立可积随机变量序列 $\xi = (\xi_n)_{n \geqslant 0}$ 是鞅 - 差.

7. 下鞅 (上鞅) 的构造 下面的定理使下鞅 (上鞅) 的结构更加清晰.

定理 2 (杜布 (J. L. Doob)) *设 $X = (X_n, \mathscr{F}_n)$ 是下鞅, 则存在鞅 $m = (m_n, \mathscr{F}_n)$ 和这样的可预测递增序列 $A = (A_n, \mathscr{F}_{n-1})$, 使对于每一个 $n \geqslant 0$, 有杜布分解:*

$$X_n = m_n + A_n \quad (\mathbf{P} - \text{a.c.}). \tag{11}$$

而且类似形式的分解唯一.

证明 设 $m_0 = X_0, A_0 = 0$, 而

$$m_n = m_0 + \sum_{j=0}^{n-1} [X_{j+1} - \mathbf{E}(X_{j+1} | \mathscr{F}_j)], \tag{12}$$

$$A_n = \sum_{j=0}^{n-1} [\mathbf{E}(X_{j+1} | \mathscr{F}_j) - X_j]. \tag{13}$$

这样定义的 m 和 A, 显然具有所要求的性质. 其次, 亦设 $X_n = m'_n + A'_n$, 其中 $m' = (m'_n, \mathscr{F}_n)$ 是鞅, 而 $A' = (A'_n, \mathscr{F}_{n-1})$ 是可预测递增序列. 那么,

$$A'_{n+1} - A'_n = (A_{n+1} - A_n) + (m_{n+1} - m_n) - (m'_{n+1} - m'_n),$$

然后在等式两侧同求条件数学期望, 得 $A'_{n+1} - A'_n = A_{n+1} - A_n$ $(\mathbf{P} - \text{a.c.})$. 由于 $A_0 = A'_0 = 0$, 说明对于一切 $n \geqslant 0, A_n = A'_n$ $(\mathbf{P} - \text{a.c.})$ 和 $m_n = m'_n$ $(\mathbf{P} - \text{a.c.})$. □

由分解 (11) 可见, 序列 $A = (A_n, \mathscr{F}_{n-1})$ 补偿 $X = (X_n, \mathscr{F}_n)$ 成为鞅. 这说明如下定义是合理的.

定义 7 杜布分解 (11) 中的可预测递增序列 $A = (A_n, \mathscr{F}_{n-1})$, 称做 (下鞅 X 的) 补偿.

如果 $\mathbf{E}M^2 < \infty, n \geqslant 0$, 则称鞅 $M = (M_n, \mathscr{F}_n)_{n \geqslant 0}$ 为平方可积的. 在研究平方可积鞅 $M = (M_n, \mathscr{F}_n)_{n \geqslant 0}$ 时, 杜布分解起核心作用, 因为根据上面的说明随机序列 $M^2 = (M_n^2, \mathscr{F}_n)$ 是下鞅. 根据定理 2, 存在鞅 $m = (m_n, \mathscr{F}_n)$ 和可预测递增序列 $\langle M \rangle = (\langle M \rangle_n, \mathscr{F}_{n-1})$, 使

$$M_n^2 = m_n + \langle M \rangle_n. \tag{14}$$

序列 $\langle M \rangle$ 称做鞅 M 的二次特征, 并在许多方面决定其构造和性质.

由 (13) 式可见,

$$\langle M \rangle_n = \sum_{j=1}^{n} \mathbf{E}[(M_j)^2 | F_{j-1}], \tag{15}$$

而且对于一切 $l \leqslant k$, 有

$$\mathbf{E}[(M_k - M_l)^2|\mathscr{F}_l] = \mathbf{E}[M_k^2 - M_l^2|\mathscr{F}_l] = \mathbf{E}[\langle M\rangle_k - \langle M\rangle_l|\mathscr{F}_l], \tag{16}$$

特别, 若 $M_0 = 0$ $(\mathbf{P}-\text{a.c.})$, 则

$$\mathbf{E}M_k^2 = \mathbf{E}\langle M\rangle_k. \tag{17}$$

有益地指出, 如果 $M_0 = 0, M_n = \xi_1 + \cdots + \xi_n$, 其中 (ξ_n) 是独立随机变量序列, 且 $\mathbf{E}\xi_i = 0, \mathbf{E}\xi_i^2 < \infty$, 则二次特征

$$\langle M\rangle_n = \mathbf{E}M_n^2 = \mathbf{D}\xi_1 + \cdots + \mathbf{D}\xi_n \tag{18}$$

是非随机的, 并且等于 M_n 的方差.

如果 $X = (X_n, \mathscr{F}_n)$ 和 $Y = (Y_n, \mathscr{F}_n)$ 是平方可积鞅, 则设

$$\langle X, Y\rangle_n = \frac{1}{4}\left[\langle X+Y\rangle_n - \langle X-Y\rangle_n\right]. \tag{19}$$

不难验证, $(X_nY_n - \langle X, Y\rangle_n, \mathscr{F}_n)$ 是鞅, 因而对于任意 $l \leqslant k$,

$$\mathbf{E}\left[(X_k - X_l)(Y_k - Y_l)|\mathscr{F}_l\right] = \mathbf{E}\left[\langle X, Y\rangle_k - \langle X, Y\rangle_l|\mathscr{F}_l\right]. \tag{20}$$

当 $X_n = \xi_1 + \cdots + \xi_n$ 和 $Y_n = \eta_1 + \cdots + \eta_n$ 时, 其中 (ξ_n) 和 (η_n) 是独立随机变量序列, 而且 $\mathbf{E}\xi_i = \mathbf{E}\eta_i = 0, \mathbf{E}\xi_i^2 < \infty, \mathbf{E}\eta_i^2 < \infty$, 变量 $\langle X, Y\rangle_n$ 等于

$$\langle X, Y\rangle_n = \sum_{i=1}^n \operatorname{cov}(\xi_i, \eta_i).$$

序列 $\langle X, Y\rangle = (\langle X, Y\rangle_n, \mathscr{F}_{n-1})$ 常称为 (平方可积) 鞅 X 和 Y 的*相互特征*. 不难验证, 有 (对照 (15) 式)

$$\langle X, Y\rangle_n = \sum_{i=1}^n \mathbf{E}[\Delta X_i \Delta Y_i|\mathscr{F}_{i-1}].$$

下面两个量在鞅论中也有重要作用: 一个是*二次协方差*

$$[X, Y]_n = \sum_{i=1}^n \Delta X_i \Delta Y_i,$$

另一个是*二次变差*

$$[X]_n = \sum_{i=1}^n (\Delta X_i)^2.$$

这两个量对于任何随机变量序列 $X = (X_n)_{n\geqslant 1}$ 和 $Y = (Y_n)_{n\geqslant 1}$ 有定义.

8. 局部鞅是鞅的条件 鉴于定理 1, 自然产生一个问题, 局部鞅 (因而, 广义鞅或鞅变换) 实际上本身也是鞅.

定理 3 1) 设随机序列 $X=(X_n,\mathscr{F}_n)_{n\geqslant 0}$ 是局部鞅 (其中 $X_0=0$, 或更一般地 $\mathbf{E}|X_0|<\infty$).

如果 $\mathbf{E}X_n^-<\infty, n\geqslant 0$, 或 $\mathbf{E}X_n^+<\infty, n\geqslant 0$, 则序列 $X=(X_n,\mathscr{F}_n)_{n\geqslant 0}$ 是鞅.

2) 设 $X=(X_n,\mathscr{F}_n)_{0\leqslant n\leqslant N}(N>0)$ 是局部鞅, 或 $\mathbf{E}X_N^-<\infty$, 或 $\mathbf{E}X_N^+<\infty$. 那么 $X=(X_n,\mathscr{F}_n)_{0\leqslant n\leqslant N}$ 是鞅.

证明 1) 证明, 若条件 "$\mathbf{E}X_n^-<\infty, n\geqslant 0$" 和 "$\mathbf{E}X_n^+<\infty, n\geqslant 0$" 中任何一个条件成立, 则条件 "$\mathbf{E}|X_n|<\infty, n\geqslant 0$" 也成立.

事实上, 假如对于一切 $n\geqslant 0, \mathbf{E}X_n^-<\infty$. 那么, 由法图 (F. Fatou) 引理, 可见

$$\begin{aligned}\mathbf{E}X_n^+ &= \mathbf{E}\varliminf_k X_{n\wedge\tau_k}^+ \leqslant \varliminf_k \mathbf{E}X_{n\wedge\tau_k}^+ = \varliminf_k[\mathbf{E}X_{n\wedge\tau_k}+\mathbf{E}X_{n\wedge\tau_k}^-]\\ &= \mathbf{E}X_0+\varliminf_k \mathbf{E}X_{n\wedge\tau_k}^- \leqslant |\mathbf{E}X_0|+\sum_{k=0}^{n}\mathbf{E}X_k^- <\infty.\end{aligned}$$

从而, $\mathbf{E}|X_n|<\infty, n\geqslant 0$.

为证明 ($\mathbf{E}(X_{n+1}|\mathscr{F}_n)=X_n, n\geqslant 0$) 的鞅性, 注意到对于任意马尔可夫时间 τ_k, 有

$$|X_{(n+1)\wedge\tau_k}|=\sum_{i=0}^{n+1}|X_i|,$$

其中

$$\mathbf{E}\sum_{i=0}^{n+1}|X_i|<\infty.$$

因此, 由勒贝格控制收敛定理, 经在关系式 $\mathbf{E}(X_{(n+1)\wedge\tau_k}|\mathscr{F}_n)=X_{n\wedge\tau_k}$ 中极限过渡 ($k\to\infty, \tau_k\uparrow\infty(\mathbf{P}-\text{a.c.})$), 得 $\mathbf{E}(X_{n+1}|\mathscr{F}_n)=X_n(\mathbf{P}-\text{a.c.})$.

2) 假设 $\mathbf{E}X_N^-<\infty$, 证明对于一切 $n<N, \mathbf{E}X_n^-<\infty$.

事实上, 由于局部鞅也是广义鞅, 可见 $X_n=\mathbf{E}(X_{n+1}|\mathscr{F}_n)$, 其中 $\mathbf{E}(|X_{n+1}||\mathscr{F}_n)<\infty$ ($\mathbf{P}-\text{a.c.}$). 那么, 根据对于条件数学期望的延森 (I. L. Iensen) 不等式 (见第二章 §7 练习题 5), 有 $X_n^-\leqslant\mathbf{E}(X_{n+1}^-|\mathscr{F}_n)$. 因此, $\mathbf{E}X_n^-\leqslant\mathbf{E}X_{n+1}^-\leqslant\mathbf{E}X_N^-<\infty$.

于是, 由命题 1), 可得所要证明的局部鞅 $X=(X_n,\mathscr{F}_n)_{0\leqslant n\leqslant N}$ 的鞅性.

9. 练习题

1. 证明条件 (2) 和 (3) 的等价性.

2. 设 σ 和 τ 是马尔可夫时间, 证明 $\tau+\sigma, \tau\vee\sigma, \tau\wedge\sigma$ 也是马尔可夫时间, 并且如果 $\sigma\leqslant\tau$, 则 $\mathscr{F}_\sigma\subseteq\mathscr{F}_\tau$.

3. 证明 τ 和 X_τ 是 $\mathscr{F}_\tau$ - 可测的.

4. 设 $Y=(Y_n,\mathscr{F}_n)$ 是鞅 (下鞅), $V=(V_n,\mathscr{F}_{n-1})$ 是可预测序列, $(V\cdot Y)_n, n\geqslant 0$ 是可积随机变量. 证明 $V\cdot Y$ 是鞅 (下鞅).

5. 设 $\mathscr{G}_1 \supseteq \mathscr{G}_2 \supseteq \cdots$ 是非增 $\sigma-$ 代数族, 而 ξ 是可积随机变量. 证明序列 $(X_n)_{n\geqslant 1}$, 其中 $X_n = \mathbf{E}(\xi|\mathscr{G}_n)$, 是逆鞅, 即对于任何 $n \geqslant 1$, 有

$$\mathbf{E}(X_n|X_{n+1}, X_{n+2}, \cdots) = X_{n+1} \quad (\mathbf{P}-\text{a.c.}).$$

6. 设 $\xi_1, \xi_2, \cdots$ 是独立随机变量, 且

$$\mathbf{P}\{\xi_i = 0\} = \mathbf{P}\{\xi_i = 2\} = \frac{1}{2} \quad 和 \quad X_n = \prod_{i=1}^{n} \xi_i.$$

证明不存在可积随机变量 ξ 和非降 $\sigma-$ 代数族 $(\mathscr{F}_n)$, 使 $X_n = \mathbf{E}(\xi|\mathscr{F}_n)$. (这个例子说明, 并不是每一个鞅 $(X_n)_{n\geqslant 1}$ 都可以表示为 $(\mathbf{E}(\xi|\mathscr{F}_n))_{n\geqslant 1}$; 对照第一章 §11 例 3).

7. (a) 设 $\xi_1, \xi_2, \cdots$ 是独立随机变量, 且对于 $n \geqslant 1, \mathbf{E}|\xi_n| < \infty, \mathbf{E}\xi_n = 0$, 证明对于每一个 $k \geqslant 1$, 序列

$$X_n^{(k)} = \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \xi_{i_1} \cdots \xi_{i_k}, \quad n \geqslant k$$

是鞅.

(b) 设 $\xi_1, \xi_2, \cdots$ 是可积随机变量, 满足

$$\mathbf{E}(\xi_{n+1}|\xi_1, \cdots, \xi_n) = \frac{\xi_1 + \cdots + \xi_n}{n} (= X_n).$$

证明序列 $X_1, X_2, \cdots$ 是鞅.

8. 举一鞅的例 $X = (X_n, \mathscr{F}_n)_{n\geqslant 1}$, 而随机变量族 $(X_n, n \geqslant 1)$ 不一致可积.

9. 设 $X = (X_n)_{n\geqslant 0}$ 是马尔可夫链 (第八章 §1), 且其状态的集合 $E = \{i, j, \cdots\}$ 是可数的, 而 p_{ij} 是它的转移概率. 此外, 假设 $\psi = \psi(x), x \in E$, 是有界函数, 并且满足条件

$$\sum_{j \in E} p_{ij}\psi(j) \leqslant \lambda\psi(i), \quad \lambda > 0, \quad i \in E.$$

证明序列 $(\lambda^{-n}\psi(X_n))_{n\geqslant 0}$ 是上鞅.

§2. 在时间变量为随机时间时鞅性的不变性

1. 杜布定理　如果 $X = (X_n, \mathscr{F}_n)_{n\geqslant 0}$ 是鞅, 则对于一切 $n \geqslant 1$,

$$\mathbf{E}X_n = \mathbf{E}X_0. \tag{1}$$

假如将时间 n 换成随机时间 (例如, 马尔可夫时间) τ, 那么上述性质是否仍然成立? 在上一节例 8 中举的例子说明, 一般并非如此: 存在鞅 X 和 (以概率 1 有限的) 马尔可夫时间 τ, 使得

$$\mathbf{E}X_\tau \neq \mathbf{E}X_0. \tag{2}$$

下面的重要定理描绘一些 "典型" 情形, 其中包括 $\mathbf{E}X_\tau = \mathbf{E}X_0$ 的情形,

定理 1 (杜布) 设 $X=(X_n,\mathscr{F}_n)$ 是鞅 (下鞅), τ_1 和 τ_2 是停止时间, 满足

$$\mathbf{E}|X_{\tau_i}|<\infty,\quad i=1,2. \tag{3}$$

$$\lim_{n\to\infty}\int_{\{\tau_2>n\}}|X_n|\mathrm{d}\mathbf{P}=0. \tag{4}$$

那么,

$$\mathbf{E}(X_{\tau_2}|\mathscr{F}_{\tau_1})\underset{(\geqslant)}{=}X_{\tau_1}\quad(\{\tau_2\geqslant\tau_1\};\mathbf{P}-\text{a.c.}). \tag{5}$$

并且, 如果 $\mathbf{P}\{\tau_1\leqslant\tau_2\}=1$, 则

$$\mathbf{E}\tau_2\underset{(\geqslant)}{=}\mathbf{E}\tau_1. \tag{6}$$

证明 只需证明, 对于任意 $A\in\mathscr{F}_{\tau_1}$, 有

$$\int_{A\cap\{\tau_2\geqslant\tau_1\}}X_{\tau_2}\mathrm{d}\mathbf{P}\underset{(\geqslant)}{=}\int_{A\cap\{\tau_2\geqslant\tau_1\}}X_{\tau_1}\mathrm{d}\mathbf{P}. \tag{7}$$

同样, 为此只需证明, 对于任意 $n\geqslant 0$, 有

$$\int_{A\cap\{\tau_2\geqslant\tau_1\}\cap\{\tau_1=n\}}X_{\tau_2}\mathrm{d}\mathbf{P}\underset{(\geqslant)}{=}\int_{A\cap\{\tau_2\geqslant\tau_1\}\cap\{\tau_1=n\}}X_{\tau_1}\mathrm{d}\mathbf{P},$$

或者

$$\int_{B\cap\{\tau_2\geqslant n\}}X_{\tau_2}\mathrm{d}\mathbf{P}\underset{(\geqslant)}{=}\int_{B\cap\{\tau_2\geqslant n\}}X_n\mathrm{d}\mathbf{P}. \tag{8}$$

其中 $B=A\bigcap\{\tau_1=n\}\in\mathscr{F}_n$.

因此, 有

$$\begin{aligned}
\int_{B\cap\{\tau_2\geqslant n\}}X_n\mathrm{d}\mathbf{P} &= \int_{B\cap\{\tau_2=n\}}X_n\mathrm{d}\mathbf{P}+\int_{B\cap\{\tau_2>n\}}X_n\mathrm{d}\mathbf{P}\\
&\underset{(\leqslant)}{=}\int_{B\cap\{\tau_2=n\}}X_n\mathrm{d}\mathbf{P}+\int_{B\cap\{\tau_2>n\}}\mathbf{E}(X_{n+1}|\mathscr{F}_n)\mathrm{d}\mathbf{P}\\
&=\int_{B\cap\{\tau_2=n\}}X_{\tau_2}\mathrm{d}\mathbf{P}+\int_{B\cap\{\tau_2\geqslant n+1\}}X_{n+1}\mathrm{d}\mathbf{P}\\
&\underset{(\leqslant)}{=}\int_{B\cap\{n\leqslant\tau_2\leqslant n+1\}}X_{\tau_2}\mathrm{d}\mathbf{P}+\int_{B\cap\{\tau_2\geqslant n+2\}}X_{n+2}\mathrm{d}\mathbf{P}\\
&\underset{(\leqslant)}{=}\int_{B\cap\{n\leqslant\tau_2\leqslant m\}}X_{\tau_2}\mathrm{d}\mathbf{P}+\int_{B\cap\{\tau_2>m\}}X_m\mathrm{d}\mathbf{P}.\\
&\underset{(\leqslant)}{=}\cdots
\end{aligned}$$

从而, 有

$$\int_{B\cap\{n\leqslant\tau_2\leqslant m\}}X_{\tau_2}\mathrm{d}\mathbf{P}\underset{(\geqslant)}{=}\int_{B\cap\{n\leqslant\tau_2\}}X_n\mathrm{d}\mathbf{P}-\int_{B\cap\{m<\tau_2\}}X_m\mathrm{d}\mathbf{P},$$

并且由于 (4) 式和等式 $X_m = 2X_m^+ - |X_m|$, 得

$$\int_{B\cap\{\tau_2\geqslant n\}} X_{\tau_2}\mathrm{d}\mathbf{P} \underset{(\geqslant)}{=} \lim_{m\to\infty}\left[\int_{B\cap\{n\leqslant\tau_2\}} X_n\mathrm{d}\mathbf{P} - \int_{B\cap\{m<\tau_2\}} X_m\mathrm{d}\mathbf{P}\right]$$
$$= \int_{B\cap\{n\leqslant\tau_2\}} X_{\tau_2}\mathrm{d}\mathbf{P} - \lim_{m\to\infty}\int_{B\cap\{m<\tau_2\}} X_m\mathrm{d}\mathbf{P}$$
$$= \int_{B\cap\{n\leqslant\tau_2\}} X_{\tau_2}\mathrm{d}\mathbf{P}.$$

于是 (8) 式得证, 故 (5) 式也得证. 最后, 由 (5) 式得 (6) 式. □

系 1　如果存在常数 N, 使 $\mathbf{P}\{\tau_1 \leqslant N\} = 1, \mathbf{P}\{\tau_2 \leqslant N\} = 1$, 则满足条件 (3),(4). 因此, 如果同时 $\mathbf{P}\{\tau_1 \leqslant \tau_2\} = 1$ 且 X 是鞅, 则

$$\mathbf{E}X_0 = \mathbf{E}\tau_1 = \mathbf{E}\tau_2 = \mathbf{E}X_N.$$

系 2　如果随机变量族 $\{X_n\}$ 一致可积 (特别, 如果以概率 1, 有 $|X_n| \leqslant C < \infty, n \geqslant 0$), 则满足条件 (3), (4).

事实上, 由于 $\mathbf{P}\{\tau_i > n\} \to 0, n \to \infty$, 故由第二章 §6 引理 2 可以得出条件 (4). 其次, 由于随机变量族 $\{X_n\}$ 一致可积, 可见 (见第二章 §6 (16) 式)

$$\sup_N \mathbf{E}|X_N| < \infty. \tag{9}$$

如果 τ 是某一停时, 而 X 是下鞅, 则将系 1 用于有界停时 $\tau_N = \tau \wedge N$, 可见

$$\mathbf{E}X_0 \leqslant \mathbf{E}X_{\tau_N}.$$

因此,

$$\mathbf{E}|X_{\tau_N}| = 2\mathbf{E}\tau_{\tau_N}^+ - \mathbf{E}X_{\tau_N} \leqslant 2\mathbf{E}\tau_{\tau_N}^+ - \mathbf{E}X_0. \tag{10}$$

序列 $X^+ = (X_n^+, \mathscr{F}_n)$ 是下鞅 (§1 例 5), 从而

$$\mathbf{E}X_{\tau_N}^+ = \sum_{j=0}^N \int_{\{\tau_N=j\}} X_j^+\mathrm{d}\mathbf{P} + \int_{\{\tau>N\}} X_N^+\mathrm{d}\mathbf{P}$$
$$\leqslant \sum_{j=0}^N \int_{\{\tau_N=j\}} X_N^+\mathrm{d}\mathbf{P} + \int_{\{\tau>N\}} X_N^+\mathrm{d}\mathbf{P} = \mathbf{E}X_N^+ \leqslant \mathbf{E}|X_N| \leqslant \sup_N \mathbf{E}|X_N|.$$

由此连同 (10) 式, 得

$$\mathbf{E}|X_{\tau_N}| \leqslant 3\sup_N \mathbf{E}|X_N|.$$

从而根据法图引理, 得

$$\mathbf{E}|X_\tau| \leqslant 3\sup_N \mathbf{E}|X_N|.$$

于是, 若选择 $\tau = \tau_i (i = 1, 2)$, 并考虑到 (9) 式, 得 $\mathbf{E}|X_{\tau_i}| < \infty, i = 1, 2$.

注 在上一节考虑的例 8 中,

$$\int_{\{\tau>n\}}|X_n|\mathrm{d}\mathbf{P}=(2^n-1)\mathbf{P}\{\tau>n\}=(2^n-1)2^{-n}\to 1,\quad n\to\infty,$$

从而, (对于 $\tau_2=\tau$) 条件 (4) 不成立.

2. 杜布定理的特例 对于应用, 由定理 1 引出的条件常是重要的.

定理 2 设 $X=(X_n)$ 是鞅 (下鞅), 而 τ 关于 $(\mathscr{F}_n^X)$ 是停时, 其中 $\mathscr{F}_n^X=\sigma\{X_0,\cdots,X_n\}$. 假设

$$\mathbf{E}\tau<\infty,$$

并且对于任意 $n\geqslant 0$ 和某个常数 $C>0$, 有

$$\mathbf{E}\{|X_{n+1}-X_n||\mathscr{F}_n^X\}\leqslant C\quad(\{\tau\geqslant n\};\mathbf{P}-\text{a.c.}).$$

那么,

$$\mathbf{E}|X_\tau|<\infty,$$

并且

$$\mathbf{E}X_\tau \underset{(\geqslant)}{=} \mathbf{E}X_0. \tag{11}$$

证明 现在验证, 对于 $\tau_2=\tau$, 定理 1 的条件 (3) 和 (4) 成立.

设 $Y_0=|X_0|, Y_j=|X_j-X_{j-1}|, j\geqslant 1$, 则

$$|X_\tau|\leqslant\sum_{j=0}^{\tau}Y_j,$$

$$\begin{aligned}\mathbf{E}|X_\tau|&\leqslant\mathbf{E}\left(\sum_{j=0}^{\tau}Y_j\right)=\int_\Omega\sum_{j=0}^{\tau}Y_j\mathrm{d}\mathbf{P}=\sum_{n=0}^{\infty}\int_{\{\tau=n\}}\sum_{j=0}^{n}Y_j\mathrm{d}\mathbf{P}\\&=\sum_{n=0}^{\infty}\sum_{j=0}^{n}\int_{\{\tau=n\}}Y_j\mathrm{d}\mathbf{P}=\sum_{j=0}^{\infty}\sum_{n=j}^{\infty}\int_{\{\tau=n\}}Y_j\mathrm{d}\mathbf{P}=\sum_{j=0}^{\infty}\int_{\{\tau\geqslant j\}}Y_j\mathrm{d}\mathbf{P},\end{aligned}$$

其中集合 $\{\tau\geqslant j\}=\Omega\backslash\{\tau<j\}\in\mathscr{F}_{j-1}^X$. 因此对于 $j\geqslant 1$, 有

$$\int_{\{\tau\geqslant j\}}Y_j\mathrm{d}\mathbf{P}=\int_{\{\tau\geqslant j\}}\mathbf{E}(Y_j|X_0,\cdots,X_{j-1})\mathrm{d}\mathbf{P}\leqslant C\mathbf{P}\{\tau\geqslant j\}.$$

于是,

$$\mathbf{E}|X_\tau|\leqslant\mathbf{E}\left(\sum_{j=0}^{\tau}Y_j\right)\leqslant C\sum_{j=1}^{\infty}\mathbf{P}\{\tau\geqslant j\}+\mathbf{E}|X_0|=C\mathbf{E}\tau+\mathbf{E}|X_0|<\infty. \tag{12}$$

其次, 如果 $\tau>n$, 则

$$\sum_{j=0}^{n}Y_j\leqslant\sum_{j=0}^{\tau}Y_j,$$

所以

$$\int_{\{\tau>n\}}|X_n|\mathrm{d}\mathbf{P}\leqslant\int_{\{\tau>n\}}\sum_{j=0}^{\tau}Y_j\mathrm{d}\mathbf{P}.$$

由此并考虑到 (根据 (12) 式), $\mathbf{E}\left(\sum\limits_{j=0}^{\tau}Y_j\right)<\infty$, 且当 $n\to\infty$ 时 $\{\tau>n\}\downarrow\varnothing$, 根据控制收敛定理, 有

$$\varliminf_{n\to\infty}\int_{\{\tau>n\}}|X_n|\mathrm{d}\mathbf{P}\leqslant\varliminf_{n\to\infty}\int_{\{\tau>n\}}\sum_{j=0}^{\tau}Y_j\mathrm{d}\mathbf{P}.$$

于是, 定理 1 的条件成立, 故由此得 (11) 式. □

3. 基本瓦尔德恒等式　现在介绍刚证明的定理的某些应用,

定理 3 (瓦尔德恒等式)　设 $\xi_1,\xi_2,\cdots$ 是独立同分布随机变量, 且 $\mathbf{E}|\xi_i|<\infty$, 而 τ 关于 $(\mathscr{F}_n^{\xi})$ 是停时, 其中 $\mathscr{F}_n^{\xi}=\sigma\{\xi_1,\cdots,\xi_n\},\tau\geqslant 1,\mathbf{E}\tau<\infty$. 那么,

$$\mathbf{E}(\xi_1+\cdots+\xi_\tau)=\mathbf{E}\xi_1\times\mathbf{E}\tau. \tag{13}$$

如果同时又 $\mathbf{E}\xi_1^2<\infty$, 则

$$\mathbf{E}[(\xi_1+\cdots+\xi_\tau)-\tau\mathbf{E}\xi_1]^2=\mathbf{D}\xi_1\times\mathbf{E}\tau. \tag{14}$$

证明　显然, $X=(X_n,\mathscr{F}_n^{\xi})_{n\geqslant 1}$ 是鞅, 其中 $X_n=(\xi_1+\cdots+\xi_\tau)-n\mathbf{E}\xi_1$, 且

$$\begin{aligned}\mathbf{E}[|X_{n+1}-X_n||X_1,\cdots,X_n]&=\mathbf{E}[|\xi_{n+1}-\mathbf{E}\xi_1||\xi_1,\cdots,\xi_n]\\&=\mathbf{E}|\xi_{n+1}-\mathbf{E}\xi_1|\leqslant 2\mathbf{E}|\xi_1|<\infty.\end{aligned}$$

因此, 根据定理 2, 得 $\mathbf{E}X_\tau=\mathbf{E}X_0=0$, 从而 (13) 式得证.

现在给出 "第二个瓦尔德恒等式" (14) 的三个证明,

证明 I　设 $\eta_i=\xi_i-\mathbf{E}\xi_i,S_n=\eta_1+\cdots+\eta_n$. 需要证明

$$\mathbf{E}S_\tau^2=\mathbf{E}\eta_1^2\times\mathbf{E}\tau.$$

设 $\tau(n)=\tau\wedge n(=\min\{\tau,n\})$. 由于

$$S_n^2=\sum_{i=1}^{n}\eta_i^2+2\sum_{1\leqslant i<j\leqslant n}\eta_i\eta_j,$$

可见序列 $\left(S_n^2-\sum\limits_{i=1}^{n}\eta_i^2,\mathscr{F}_n^{\xi}\right)_{n\geqslant 1}$ 是均值为 0 的鞅.

由定理 1 的系 1, 可见

$$\mathbf{E}S_{\tau(n)}^2=\mathbf{E}\sum_{i=1}^{\tau(n)}\eta_i^2.$$

根据“第一个瓦尔德恒等式” (13), 有

$$\mathbf{E}\sum_{i=1}^{\tau(n)}\eta_i^2=\mathbf{E}\eta_1^2\times\mathbf{E}\tau(n),$$

因而有 $\mathbf{E}S_\tau^2=\mathbf{E}\eta_1^2\times\mathbf{E}\tau$.

由于根据假设当 $m,n\to\infty$ 时, 类似地可得

$$\mathbf{E}(S_{\tau(n)}-S_{\tau(m)})^2=\mathbf{E}\eta_1^2\times\mathbf{E}[\tau(n)-\tau(m)]\to 0.$$

于是, 序列 $\{S_{\tau(n)}\}_{n\geqslant 1}$ 是 L^2 中的基本序列 (柯西序列, 见第二章 §10 第 5 小节), 从而根据第二章 §10 定理 7, 存在随机变量 S, 使 $\mathbf{E}(S_{\tau(n)}-S)^2\to 0, n\to\infty$. 由此可见 (第二章 §11 练习题 1), $\mathbf{E}S_{\tau(n)}^2\to\mathbf{E}S^2, n\to\infty$. 上面已证明 $\mathbf{E}S_{\tau(n)}^2=\mathbf{E}\eta_1^2\times\mathbf{E}\tau(n)$, 故当 $n\to\infty$ 时得 $\mathbf{E}S^2=\mathbf{E}\eta_1^2\times\mathbf{E}\tau$.

现在只剩下识别变量 S. 为此只需注意到存在一序列 $\{n'\}\subseteq\{n\}$, 使以概率 1 同时有收敛: $S_{\tau(n')}\to S$ 和 $\tau(n')\to\tau$. 那么, 以概率 1 也有收敛 $S_{\tau(n')}\to S_\tau$. 从而, 以概率 1 有 $S=S_\tau$, 故 $\mathbf{E}S_\tau^2=\mathbf{E}\eta_1^2\times\mathbf{E}\tau$, 而这正是要证明的.

证明 Ⅱ 由已证明的等式 $\mathbf{E}S_{\tau(n)}^2=\mathbf{E}\eta_1^2\times\mathbf{E}\tau(n)$ 和法图引理 (见第二章 §6 定理 2 中的 a)), 可见

$$\mathbf{E}S_\tau^2=\mathbf{E}\underline{\lim}S_{\tau(n)}^2\leqslant\underline{\lim}\mathbf{E}S_{\tau(n)}^2\leqslant\mathbf{E}\eta_1^2\times\mathbf{E}\tau.$$

如果能证明对所有 $n\geqslant 1$, 不等式 $\mathbf{E}S_{\tau(n)}^2\leqslant\mathbf{E}S_\tau^2$ 成立, 则欲证的等式 $\mathbf{E}S_\tau^2=\mathbf{E}\eta_1^2\times\mathbf{E}\tau$ 就随之得证.

为此注意到, 由“第一个瓦尔德恒等式” (13), 有

$$\mathbf{E}|S_\tau|=\mathbf{E}|\eta_1+\cdots+\eta_\tau|\leqslant\mathbf{E}(|\eta_1|+\cdots+|\eta_\tau|)=\mathbf{E}|\eta_1|\times\mathbf{E}\tau<\infty.$$

因而, 当 $n\to\infty$ 时, 有

$$\begin{aligned}&\mathbf{E}|S_n|I(\tau>n)=\mathbf{E}|\eta_1+\cdots+\eta_n|I(\tau>n)\\ \leqslant&\ \mathbf{E}(|\eta_1|+\cdots+|\eta_n|)I(\tau>n)\leqslant\mathbf{E}(|\eta_1|+\cdots+|\eta_\tau|)I(\tau>n)\to 0.\end{aligned}$$

运用定理 1 (设 $\tau_1=n,\tau_2=\tau$ 和下鞅 $(|S_n|,\mathscr{F}_n^\xi)_{n\geqslant 1}$) 可得, 在集合 $\{\tau\geqslant n\}$ 上, 以概率 1 有

$$\mathbf{E}(|S_\tau||\mathscr{F}_n^\xi)\geqslant|S_n|.$$

由 (关于条件数学期望得) 延森不等式 (第二章 §7 练习题 5) 可见, 在集合 $\{\tau\geqslant n\}$ 上, 以概率 1 有

$$\mathbf{E}(S_\tau^2|\mathscr{F}_n^\xi)\geqslant S_n^2=S_{\tau(n)}^2.$$

但在集合 $\{\tau < n\}$ 上 $\mathbf{E}(S_\tau^2|\mathscr{F}_n^\xi) = S_\tau^2 = S_{\tau(n)}^2$. 所以以概率 1 有

$$\mathbf{E}(S_\tau^2|\mathscr{F}_n^\xi) \geqslant S_{\tau(n)}^2.$$

因此, $\mathbf{E}S_\tau^2 \geqslant S_{\tau(n)}^2$. 而这正是要证明的.

证明 Ⅲ 由 "第一个证明" 可见 $\left(S_n^2 - \sum_{i=1}^n \eta_i^2, \mathscr{F}_n^\xi\right)_{n\geqslant 1}$ 是鞅, 而对于 $\tau(n) = \tau \wedge n$, 有

$$\mathbf{E}S_{\tau(n)}^2 = \mathbf{E}\eta_1^2 \times \mathbf{E}\tau(n).$$

因为 $\mathbf{E}\tau(n) \to \mathbf{E}\tau$, 故只需证明 $\mathbf{E}S_{\tau(n)}^2 \to \mathbf{E}S_\tau^2$. 而为此只需证明

$$\mathbf{E}\sup_n S_{\tau(n)}^2 < \infty,$$

因为, 这时由勒贝格控制收敛定理 (第二章 §6 定理 3) 可得要求证明的收敛性.

为了证明不等式 $\mathbf{E}\sup\limits_n S_{\tau(n)}^2 < \infty$, 我们利用将在下面 §3 中引进的 "最大不等式" (14). 若将该不等式用于鞅 $(S_{\tau(k)}, \mathscr{F}_k^\xi)_{k\geqslant 1}$, 则得

$$\mathbf{E}\left[\sup_{1\leqslant k\leqslant n} S_{\tau(k)}^2\right] \leqslant 4\mathbf{E}S_{\tau(n)}^2 \leqslant 4\sup_n \mathbf{E}S_{\tau(n)}^2,$$

由此利用单调收敛定理 (第二章 §6 定理 1), 得

$$\mathbf{E}\sup_{k\geqslant 1} S_{\tau(k)}^2 \leqslant 4\sup_n \mathbf{E}S_{\tau(n)}^2.$$

因为

$$\mathbf{E}S_{\tau(n)}^2 = \mathbf{E}\eta_1^2 \times \mathbf{E}\tau(n) \leqslant \mathbf{E}\eta_1^2 \times \mathbf{E}\tau < \infty,$$

所以

$$\mathbf{E}\sup_n S_{\tau(n)}^2 \leqslant 4\mathbf{E}\eta_1^2 \times \mathbf{E}\tau < \infty,$$

而这正是需要证明的. □

系 设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量序列, 且 $\mathbf{P}\{\xi_i = 1\} = \mathbf{P}\{\xi_i = -1\} = 1/2, S_n = \xi_1 + \cdots + \xi_n$, 而 $\tau = \inf\{n \geqslant 1 : S_n = 1\}$. 那么, $\mathbf{P}\{\tau < \infty\} = 1$ (例如, 参见第一章 §9 的 (20) 式), 因此 $\mathbf{P}\{S_\tau = 1\} = 1, \mathbf{E}S_\tau = 1$. 从而, 由 (13) 式可见 $\mathbf{E}\tau = \infty$.

定理 4 (瓦尔德基本恒等式) 设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量序列, 且 $S_n = \xi_1 + \cdots + \xi_n, n \geqslant 1$. 假设 $\varphi(t) = \mathbf{E}e^{t\xi_1}, t \in \mathbb{R}$, 而且对于某个 $t_0 \neq 0, \varphi(t_0)$ 存在, 并且 $\varphi(t_0) \geqslant 1$.

如果 $\tau (\tau \geqslant 1)$ (关于 $\mathscr{F}_n^\xi = \sigma(\xi_1, \cdots, \xi_n), n \geqslant 1$) 是停时, 使 $|S_n| \leqslant C(\{\tau \geqslant n\}; \mathbf{P}-\text{a.c.})$, 而且 $\mathbf{E}\tau < \infty$, 则

$$\mathbf{E}\left[\frac{e^{t_0 S_\tau}}{[\varphi(t_0)]^\tau}\right] = 1. \tag{15}$$

证明 记

$$Y_n = e^{t_0 S_n}[\varphi(t_0)]^{-n}.$$

那么, $Y = (Y_n, \mathscr{F}_n^{\xi})_{n\geqslant 1}$ 是鞅, 其中 $\mathbf{E}Y_n = 1$, 而在集合 $\{\tau = n\}$ 上, 有

$$\begin{aligned}\mathbf{E}[|Y_{n+1} - Y_n| | Y_1, \cdots, Y_n] &= Y_n \mathbf{E}\left\{\left|\frac{e^{t_0\xi_{n+1}}}{\varphi(t_0)} - 1\right| \Big| \xi_1, \cdots, \xi_n\right\} \\ &= Y_n \mathbf{E}\left|e^{t_0\xi_1}[\varphi(t_0)]^{-1} - 1\right| \leqslant B < \infty,\end{aligned}$$

其中 B 是某一常数. 于是, 由于 $\mathbf{E}Y_n = 1$, 故由定理 2, 得 (15) 式. □

例 1 这个例子在于演示, 将上述结果用于博弈问题, 包括求破产概率和平均博弈时间 (见第一章 §9).

设 $\xi_1, \xi_2, \cdots$ 是独立伯努利随机变量序列: $\mathbf{P}\{\xi_i = 1\} = p, \mathbf{P}\{\xi_i = -1\} = q, p+q = 1$, 而 $S_n = \xi_1 + \cdots + \xi_n$, 且

$$\tau = \inf\{n \geqslant 1 : S_n = B \text{ 或 } A\}, \tag{16}$$

其中 $(-A)$ 和 B 是正整数.

由第一章 §9 的 (20) 式可见, $\mathbf{P}\{\tau < \infty\} = 1$ 和 $\mathbf{E}\tau < \infty$. 那么, 如果 $\alpha = \mathbf{P}\{S_\tau = A\}, \beta = \mathbf{P}\{S_\tau = B\}$, 则 $\alpha + \beta = 1$, 并且当 $p = q = 1/2$ 时, 由 (13) 式, 可见

$$0 = \mathbf{E}S_\tau = \alpha A + \beta B,$$

因此

$$\alpha = \frac{B}{B + |A|}, \quad \beta = \frac{|A|}{B + |A|}.$$

利用 (14) 式, 得

$$\mathbf{E}\tau = \mathbf{E}S_\tau^2 = \alpha A^2 + \beta B^2 = |AB|.$$

假如 $p \neq q$, 那么, 若考虑 $((q/p)^{S_n})_{n\geqslant 1}$ 鞅, 则可得

$$\mathbf{E}\left(\frac{q}{p}\right)^{S_\tau} = \mathbf{E}\left(\frac{q}{p}\right)^{S_1} = 1,$$

因此

$$\alpha\left(\frac{q}{p}\right)^A + \beta\left(\frac{q}{p}\right)^B = 1.$$

由此连同等式 $\alpha + \beta = 1$, 得

$$\alpha = \frac{\left(\frac{q}{p}\right)^B - 1}{\left(\frac{q}{p}\right)^B - \left(\frac{q}{p}\right)^A}, \quad \beta = \frac{1 - \left(\frac{q}{p}\right)^A}{\left(\frac{q}{p}\right)^B - \left(\frac{q}{p}\right)^A}. \tag{17}$$

注意到 $\mathbf{E}S_\tau=(p-q)\mathbf{E}\tau$, 最后得

$$\mathbf{E}\tau=\frac{\mathbf{E}S_\tau}{p-q}=\frac{\alpha A+\beta B}{p-q},$$

其中 α 和 β 由 (17) 式决定.

例 2 假设在例 1 中 $p=q=1/2$. 证明, 对于任意 $\lambda, 0<\lambda<\pi/(B+|A|)$ 和由 (16) 式决定的停时 τ,

$$\mathbf{E}(\cos\lambda)^{-\tau}=\frac{\cos\left(\lambda\dfrac{B+A}{2}\right)}{\cos\left(\lambda\dfrac{B+|A|}{2}\right)}. \tag{18}$$

为此, 考虑鞅 $X=(X_n,\mathscr{F}_n^\xi)_{n\geqslant 0}$, 其中

$$X_n=(\cos\lambda)^{-n}\cos\left[\lambda\left(S_n-\frac{B+A}{2}\right)\right], \tag{19}$$

而 $S_0=0$. 显然

$$\mathbf{E}X_n=\mathbf{E}X_0=\mathbf{E}\cos\left(\lambda\frac{B+A}{2}\right), \tag{20}$$

现在证明, 变量族 $\{X_{n\wedge\tau}\}$ 一致可积. 为此注意到, 由于定理 1 的系 1, 对于任意 $\lambda, 0<\lambda<\pi/(B+|A|)$, 有

$$\begin{aligned}\mathbf{E}X_0=\mathbf{E}X_{n\wedge\tau}&=\mathbf{E}(\cos\lambda)^{-(n\wedge\tau)}\cos\left[\lambda\left(S_{n\wedge\tau}-\frac{B+A}{2}\right)\right]\\&\geqslant\mathbf{E}(\cos\lambda)^{-(n\wedge\tau)}\cos\left(\lambda\frac{B-A}{2}\right).\end{aligned}$$

因此, 由 (20) 式, 可见

$$\mathbf{E}(\cos\lambda)^{-(n\wedge\tau)}\leqslant\frac{\cos\left(\lambda\dfrac{B+A}{2}\right)}{\cos\left(\lambda\dfrac{B+|A|}{2}\right)},$$

故根据法图引理, 有

$$\mathbf{E}(\cos\lambda)^{-\tau}\leqslant\frac{\cos\left(\lambda\dfrac{B+A}{2}\right)}{\cos\left(\lambda\dfrac{B+|A|}{2}\right)}. \tag{21}$$

从而, 根据 (19) 式

$$|X_{n\wedge\tau}|=(\cos\lambda)^{-\tau},$$

故连同 (21) 式便证明了变量族 $\{X_{n\wedge\tau}\}$ 一致可积. 那么, 由定理 1 的系 2, 可见

$$\cos\left(\lambda\frac{B+A}{2}\right)=\mathbf{E}X_0=\mathbf{E}X_\tau=\mathbf{E}(\cos\lambda)^{-\tau}\cos\left(\lambda\frac{B-A}{2}\right),$$

于是, 等式 (18) 得证.

4. 更新理论的基本定理 作为瓦尔德恒等式 (13) 的一种应用, 我们证明更新理论的所谓基本定理: 如果 $N = (N_t)_{t\geqslant 0}$ 是更新过程, 其中

$$N_t = \sum_{n=1}^{\infty} I(T_n \leqslant t), \quad T_n = \sigma_1 + \cdots + \sigma_n,$$

而 $\sigma_1, \sigma_2, \cdots$ 是独立同分布正值随机变量序列 (见第二章 §9 第 4 小节), $\mu = \mathbf{E}\sigma_1 < \infty$, 则更新函数 $m(t) = \mathbf{E}N_t$ 具有如下性质:

$$\frac{m(t)}{t} \to \frac{1}{\mu}, \quad t \to \infty. \tag{22}$$

回忆, 过程 $N = (N_t)_{t\geqslant 0}$ 本身服从强大数定律:

$$\frac{N_t}{t} \to \frac{1}{\mu}(\mathrm{P} - \mathrm{a.c.}), \quad t \to \infty.$$

(例如, 参见第四章 §3 的例 4.)

为证明 (22) 式, 只需证明

$$\varliminf_{t\to\infty} \frac{m(t)}{t} \geqslant \frac{1}{\mu} \quad 和 \quad \varlimsup_{t\to\infty} \frac{m(t)}{t} \leqslant \frac{1}{\mu}. \tag{23}$$

为此, 注意到

$$T_{N_t} \leqslant t < T_{N_t+1}. \quad t > 0. \tag{24}$$

因为对于任意 $n \geqslant 1$,

$$\{N_t + 1 \leqslant n\} = \{N_t \leqslant n-1\} = \{N_t < n\} = \{T_n > t\} = \left\{\sum_{k=1}^{n} \sigma_k > t\right\} \in \mathscr{F}_n,$$

其中 $\mathscr{F}_n$ 是由变量 $\sigma_1, \sigma_2, \cdots$ 诱导的 $\sigma-$ 代数, $\mu = \mathbf{E}\sigma_1 < \infty$, 则 (对于每一个固定的 $t > 0$) 时刻 $N_t + 1$ (但不是 N_t) 是马尔可夫时间. 那么, 由瓦尔德恒等式 (13), 可以得到

$$\mathbf{E}T_{N_t+1} = \mu[m(t) + 1], \tag{25}$$

因此, 由 (24) 式的第一个不等式, 得

$$t < \mu[m(t) + 1], \quad 即 \quad \frac{m(t)}{t} > \frac{1}{\mu} - \frac{1}{t}, \tag{26}$$

故当 $t \to \infty$ 时得 (23) 式的第一个不等式.

其次, 由不等式 (24) 的右侧可见, $t \geqslant \mathbf{E}T_{N_t}$. 由于 $T_{N_t+1} = T_{N_t} + \sigma_{N_t+1}$, 则

$$t \geqslant \mathbf{E}T_{N_t} = \mathbf{E}(T_{N_t+1} - \sigma_{N_t+1}) = \mu[m(t) + 1] - \mathbf{E}\sigma_{N_t+1}. \tag{27}$$

如果假设变量 σ_i 都有上界 ($\sigma_i \leqslant c$), 则由 (27) 得 $t \geqslant \mu[m(t)+1]-c$, 因而

$$\frac{m(t)}{t} \leqslant \frac{1}{\mu} - \frac{1}{t}\cdot\frac{c-\mu}{\mu}. \tag{28}$$

那么, 由此得 (23) 式得第二个不等式.

为去掉限制 $\sigma_i \leqslant c, i \geqslant 1$, 对于某一 $c>0$, 引进变量

$$\sigma_i^c = \sigma_i I(\sigma_i < c) + cI(\sigma_i \geqslant c),$$

并且将其与更新过程 $N^c = (N_t^c)_{t\geqslant 0}$ 相联系, 其中

$$N_t^c = \sum_{n=1}^{\infty} I(T_n^c \leqslant t), \quad T_n^c = \sigma_1^c + \cdots + \sigma_n^c.$$

由于 $\sigma_i^c \leqslant \sigma_i (i \geqslant 1)$, 则 $N_t^c \geqslant N_t$, 因而 $m^c(t) = \mathbf{E}N_t^c \geqslant \mathbf{E}N_t = m(t)$. 那么, 由 (28) 式可见

$$\frac{m(t)}{t} \leqslant \frac{m^c(t)}{t} \leqslant \frac{1}{\mu^c} + \frac{1}{t}\cdot\frac{c-\mu^c}{\mu^c}.$$

其中 $\mu^c = \mathbf{E}\sigma_1^c$.

从而

$$\varlimsup_{t\to\infty} \frac{m(t)}{t} \leqslant \frac{1}{\mu^c}.$$

现在设 $c\to\infty$, 并考虑到当 $c\to\infty$ 时 $\mu^c\to\mu$, 由此得 (23) 式得第二个不等式.

于是, 性质 (22) 得证.

注　关于更新理论更一般的结果, 例如, 可以参见 [7, 第 9 章],[69, 卷 1, 第 XIII 章].

5. 练习题

1. 对于下鞅的情形, 证明, 如果将条件 (4) 换成条件

$$\varliminf_{n\to\infty} \int_{\{\tau_2>n\}} X_n^+ \mathrm{d}\mathbf{P} = 0,$$

则定理 1 仍然成立.

2. 设 $X=(X_n,\mathscr{F}_n)_{n\geqslant 0}$ 是平方可积鞅, τ 是停时, $\mathbf{E}X_0=0$,

$$\varliminf_{n\to\infty} \int_{\{\tau_2>n\}} X_n^2 \mathrm{d}\mathbf{P} = 0.$$

证明

$$\mathbf{E}X_\tau^2 = \mathbf{E}\langle X\rangle_\tau \quad \left(= \mathbf{E}\sum_{j=0}^{\tau}(\Delta X_j)^2\right),$$

其中 $\Delta X_0 = X_0, \Delta X_j = X_j - X_{j-1}, j\geqslant 1$.

3. 证明, 对于每一个鞅 $X = (X_n, \mathscr{F}_n)_{n\geqslant 0}$ 或非负下鞅, 以及停时 τ, 有

$$\mathbf{E}|X_\tau| \leqslant \varliminf_{n\to\infty} \mathbf{E}|X_n|.$$

4. 设 $X = (X_n, \mathscr{F}_n)_{n\geqslant 0}$ 是这样的上鞅, 使 $X_n \geqslant \mathbf{E}(\xi|\mathscr{F}_n)(\mathbf{P}-\text{a.c.}), n \geqslant 0$, 其中 $\mathbf{E}|\xi| < \infty$. 证明, 如果 τ_1 和 τ_2 是停时, 且 $\mathbf{P}\{\tau_1 \leqslant \tau_2\} = 1$, 则

$$X_{\tau_1} \geqslant \mathbf{E}(X_{\tau_2}|\mathscr{F}_{\tau_1}) \ \ (\mathbf{P}-\text{a.c.}).$$

5. 设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量序列, 且 $\mathbf{P}\{\xi_i = 1\} = \mathbf{P}\{\xi_i = -1\} = 1/2$, 而 a 和 $b(b > a)$ 是正数,

$$X_n = a\sum_{k=1}^{n} I(\xi_k = +1) - b\sum_{k=1}^{n} I(\xi_k = -1),$$

而

$$\tau = \inf\{n \geqslant 1 : X_n \leqslant -r\}, \quad r > 0.$$

证明: 当 $\lambda \leqslant \alpha_0$ 时, $\mathbf{E}e^{\lambda\tau} < \infty$, 而当 $\lambda > \alpha_0$ 时, $\mathbf{E}e^{\lambda\tau} = \infty$, 其中

$$\alpha_0 = \frac{b}{a+b}\ln\frac{2b}{a+b} + \frac{a}{a+b}\ln\frac{2a}{a+b}.$$

6. 设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量序列, 且 $\mathbf{E}\xi_i = 0, \mathbf{D}\xi_i = \sigma_i^2, S_n = \xi_1 + \cdots + \xi_n, \mathscr{F}_n^\xi = \sigma(\xi_1, \cdots, \xi_n)$. 对于推广瓦尔德恒等式 (13) 和 (14), 证明如下论断的正确性:

1) 如果 $\mathbf{E}\sum_{j=1}^{\tau}\mathbf{E}|\xi_j| < \infty$, 则 $\mathbf{E}S_\tau = 0$;

2) 如果 $\mathbf{E}\sum_{j=1}^{\tau}\mathbf{E}\xi_j^2 < \infty$, 则

$$\mathbf{E}S_\tau^2 = \mathbf{E}\sum_{j=1}^{\tau}\xi_j^2 = \mathbf{E}\sum_{j=1}^{\tau}\sigma_j^2. \tag{29}$$

7. 设 $X = (X_n, \mathscr{F}_n)_{n\geqslant 1}$ 是平方可积鞅, 而 τ 是停时. 证明

$$\mathbf{E}X_\tau^2 \leqslant \mathbf{E}\sum_{n=1}^{\tau}(\Delta X_n)^2.$$

证明, 如果

$$\varliminf_{n\to\infty}\mathbf{E}[X_n^2 I(\tau > n)] < \infty \quad \text{或} \quad \varliminf_{n\to\infty}\mathbf{E}[|X_n| I(\tau > n)] = 0,$$

则

$$\mathbf{E}(\Delta X_\tau)^2 = \mathbf{E}\sum_{n=1}^{\tau}X_n^2.$$

8. 设 $X = (X_n, \mathscr{F}_n)_{n\geqslant 1}$ 是下鞅, 而对于停时 $\tau_1 \leqslant \tau_2 \leqslant \cdots$, $\mathbf{E}X_{\tau_m}$ 有定义. 证明

$$\lim_{n\to\infty} \mathbf{E}[X_n^+(\tau_m > n)] = 0, \quad m \geqslant 1.$$

证明序列 $(X_{\tau_m}, \mathscr{F}_{\tau_m})_{m\geqslant 1}$ 是下鞅 (通常 $\mathscr{F}_{\tau_m} = \{A \in \mathscr{F} : A \bigcap \{\tau_m = j\} \in \mathscr{F}_j, j \geqslant 1\}$).

§3. 一些基本不等式

1. 概率的最大不等式和 L^p 中的最大不等式　如果 $X = (X_n, \mathscr{F}_n)_{n\geqslant 0}$ 是随机序列,

$$X_n^* = \max_{0\leqslant j\leqslant n} |X_j|, \quad \|X_n\|_p = (\mathbf{E}|X_n|^p)^{1/p}, \quad p > 0.$$

下面是属于杜布的三个定理: 定理 1 ～定理 3 . 这些定理, 对于下鞅、上鞅和鞅, 给出了基本的 "概率的最大不等式" 和 " L^p 中的最大不等式".

定理 1　I. 设 $X = (X_n, \mathscr{F}_n)_{n\geqslant 0}$ 是下鞅, 则对于任意 $\lambda > 0$, 有

$$\lambda \mathbf{P}\left\{\max_{k\leqslant n} X_k \geqslant \lambda\right\} \leqslant \mathbf{E}\left[X_n^+ I\left(\max_{k\leqslant n} X_k \geqslant \lambda\right)\right] \leqslant \mathbf{E}X_n^+, \tag{1}$$

$$\lambda \mathbf{P}\left\{\min_{k\leqslant n} X_k \leqslant -\lambda\right\} \leqslant \mathbf{E}\left[X_n I\left(\min_{k\leqslant n} X_k > -\lambda\right)\right] - \mathbf{E}X_0 \leqslant \mathbf{E}X_n^+ - \mathbf{E}X_0, \tag{2}$$

$$\lambda \mathbf{P}\left\{\max_{k\leqslant n} |X_k| \geqslant \lambda\right\} \leqslant 3 \max_{k\leqslant n} \mathbf{E}|X_k|. \tag{3}$$

II. 设 $Y = (Y_n, \mathscr{F}_n)_{n\geqslant 0}$ 是上鞅, 则对于任意 $\lambda > 0$, 有

$$\lambda \mathbf{P}\left\{\max_{k\leqslant n} Y_k \geqslant \lambda\right\} \leqslant \mathbf{E}Y_0 - \mathbf{E}\left[Y_n I\left(\max_{k\leqslant n} Y_k < \lambda\right)\right] \leqslant \mathbf{E}Y_0 + \mathbf{E}Y_n^-, \tag{4}$$

$$\lambda \mathbf{P}\left\{\min_{k\leqslant n} Y_k \leqslant -\lambda\right\} \leqslant -\mathbf{E}\left[Y_n I\left(\min_{k\leqslant n} Y_k \leqslant -\lambda\right)\right] \leqslant \mathbf{E}Y_n^- \tag{5}$$

$$\lambda \mathbf{P}\left\{\max_{k\leqslant n} |Y_k| \geqslant \lambda\right\} \leqslant 3 \max_{k\leqslant n} \mathbf{E}|Y_k|. \tag{6}$$

III. 设 $X = (X_n, \mathscr{F}_n)_{n\geqslant 0}$ 是非负上鞅, 则对于任意 $\lambda > 0$, 有

$$\lambda \mathbf{P}\left\{\max_{k\leqslant n} Y_k \geqslant \lambda\right\} \leqslant \mathbf{E}Y_0, \tag{7}$$

$$\lambda \mathbf{P}\left\{\sup_{k\geqslant n} Y_k \geqslant \lambda\right\} \leqslant \mathbf{E}Y_n. \tag{8}$$

定理 2　设 $X = (X_n, \mathscr{F}_n)_{n\geqslant 0}$ 是非负下鞅, 则对于任意 $p \geqslant 1$, 下列不等式成立:

如果 $p>1$, 则

$$\|X_n\|_p \leqslant \|X_n^*\|_p \leqslant \frac{p}{p-1}\|X_n\|_p; \tag{9}$$

如果 $p=1$, 则

$$\|X_n\|_1 \leqslant \|X_n^*\|_1 \leqslant \frac{\mathrm{e}}{\mathrm{e}-1}\{1+\|X_n \ln^+ X_n\|_1\}. \tag{10}$$

定理 3 设 $X=(X_n,\mathscr{F}_n)_{n\geqslant 0}$ 是鞅, 而 $\lambda>0, p\geqslant 1$, 则

$$\mathbf{P}\left\{\max_{k\leqslant n}|X_k|\geqslant\lambda\right\}\leqslant\frac{\mathbf{E}|X_n|^p}{\lambda^p}. \tag{11}$$

而如果 $p>1$, 则

$$\|X_n\|_p \leqslant \|X_n^*\|_p \leqslant \frac{p}{p-1}\|X_n\|_p. \tag{12}$$

特别, 当 $p=2$ 时, 有

$$\mathbf{P}\left\{\max_{k\leqslant n}|X_k|\geqslant\lambda\right\}\leqslant\frac{\mathbf{E}|X_n|^2}{\lambda^2}, \tag{13}$$

$$\mathbf{E}\left[\max_{k\leqslant n}X_k^2\right]\leqslant 4\mathbf{E}|X_n|^2. \tag{14}$$

证明定理 1 由于带相反符号的下鞅是上鞅, 故由不等式 (4)~(6) 得不等式 (1)~(3). 因此, 我们考虑上鞅 $Y=(Y_n,\mathscr{F}_n)_{n\geqslant 0}$ 的情形.

设 $\tau=\inf\{k\leqslant n: Y_k\geqslant\lambda\}$, 如果 $\max\limits_{k\leqslant n}Y_k<\lambda$, 则认为 $\tau=n$. 那么, 由 §2 性质 (6) 可见,

$$\begin{aligned}\mathbf{E}Y_0\geqslant\mathbf{E}Y_\tau&=\mathbf{E}\left[Y_\tau;\max_{k\leqslant n}Y_k\geqslant\lambda\right]+\mathbf{E}\left[Y_\tau;\max_{k\leqslant n}Y_k<\lambda\right]\\&\geqslant\lambda\mathbf{P}\left[\max_{k\leqslant n}Y_k\geqslant\lambda\right]+\mathbf{E}\left[Y_n;\max_{k\leqslant n}Y_k<\lambda\right],\end{aligned}$$

从而, 证明了 (4) 式.

现在, 设 $\sigma=\inf\{k\leqslant n: Y_k\leqslant-\lambda\}$, 如果 $\max\limits_{k\leqslant n}Y_k>-\lambda$, 则认为 $\sigma=n$. 那么, 仍然由 §2 性质 (6) 可见,

$$\begin{aligned}\mathbf{E}Y_n\leqslant\mathbf{E}Y_\tau&=\mathbf{E}\left[Y_\tau;\min_{k\leqslant n}Y_k\leqslant-\lambda\right]+\mathbf{E}\left[Y_\tau;\min_{k\leqslant n}Y_k>-\lambda\right]\\&\leqslant-\lambda\mathbf{P}\left[\min_{k\leqslant n}Y_k\leqslant-\lambda\right]+\mathbf{E}\left[Y_n;\min_{k\leqslant n}Y_k>-\lambda\right].\end{aligned}$$

由此, 得

$$\lambda\mathbf{P}\left[\min_{k\leqslant n}Y_k\leqslant-\lambda\right]\leqslant-\mathbf{E}\left[Y_n;\min_{k\leqslant n}Y_k\leqslant-\lambda\right]\leqslant\mathbf{E}Y_n^-,$$

从而, 证明了 (5) 式.

为证明不等式 (6), 注意到 $Y^- = (-Y)^+$ 是下鞅, 从而根据 (4) 式和 (1) 式, 有

$$\begin{aligned}\lambda\mathbf{P}\left[\max_{k\leqslant n}|Y_k|\geqslant\lambda\right]&\leqslant\lambda\mathbf{P}\left[\max_{k\leqslant n}Y_k^+\geqslant\lambda\right]+\lambda\mathbf{P}\left[\max_{k\leqslant n}Y_k^-\geqslant\lambda\right]\\&=\lambda\mathbf{P}\left\{\max_{k\leqslant n}Y_k\geqslant\lambda\right\}+\lambda\mathbf{P}\left\{\max_{k\leqslant n}Y_k^-\geqslant\lambda\right\}\\&\leqslant\mathbf{E}Y_0+2\mathbf{E}Y_n^-\leqslant 3\max_{k\leqslant n}\mathbf{E}|Y_k|.\end{aligned}$$

由 (4) 式得不等式 (7).

为证明不等式 (8), 如果当 $k\geqslant n$ 时 $Y_k<\lambda$, 则设 $\gamma=\inf\{k\geqslant n: Y_k\geqslant\lambda\}$, 认为 $\gamma=\infty$. 又设 $n<N<\infty$. 那么, 由 §2 (6) 式, 有

$$\mathbf{E}Y_n\geqslant\mathbf{E}Y_{\gamma\wedge N}\geqslant\mathbf{E}[Y_{\gamma\wedge N}I(\gamma\leqslant N)]\geqslant\lambda\mathbf{P}\{\gamma\leqslant N\},$$

于是, 当 $N\to\infty$ 时, 得

$$\mathbf{E}Y_n\geqslant\lambda\mathbf{P}\{\gamma<\infty\}=\lambda\mathbf{P}\left\{\sup_{k\geqslant n}Y_k\geqslant n\right\}.\qquad\square$$

证明定理 2 (9) 式和 (10) 式的前两个不等式显然.

为证明 (9) 式的后两个不等式, 先假设

$$\|X_n^*\|_p<\infty,\tag{15}$$

并且利用如下事实: 对于任意非负随机变量 ξ 和 $r>0$, 有

$$\mathbf{E}\xi^r=r\int_0^\infty t^{r-1}\mathbf{P}\{\xi\geqslant t\}\mathrm{d}t.\tag{16}$$

那么, 对于 $p>1$, 由不等式 (1) 和傅比尼 (G. Fubini) 定理, 可见

$$\begin{aligned}\mathbf{E}(X_n^*)^p&=p\int_0^\infty t^{p-1}\mathbf{P}\{X_n^*\geqslant t\}\mathrm{d}t\leqslant p\int_0^\infty t^{p-2}\left(\int_{\{X_n^*\geqslant t\}}X_n\mathrm{d}\mathbf{P}\right)\mathrm{d}t\\&=p\int_0^\infty t^{p-2}\left[\int_\Omega X_nI\{X_n^*\geqslant t\}\mathrm{d}\mathbf{P}\right]\mathrm{d}t=p\int_\Omega X_n\left[\int_0^{X_n^*}t^{p-2}\mathrm{d}t\right]\mathrm{d}\mathbf{P}\\&=\frac{p}{p-1}\mathbf{E}[X_n(X_n^*)^{p-1}].\end{aligned}\tag{17}$$

从而, 根据赫尔德 (O. L. Hölder) 不等式 (第二章 §6 第 7 小节), 有

$$\mathbf{E}(X_n^*)^p\leqslant q\|X_n\|_p\cdot\|(X_n^*)^{p-1}\|_q=q\|X_n\|_p[\mathbf{E}(X_n^*)^p]^{1/q},\tag{18}$$

其中 $q=p/(p-1)$.

假如不等式 (15) 成立, 则由 (18) 式立即得到 (9) 式的第二个不等式.

假如不等式 (15) 不成立, 则应该按如下方式操作. 在 (17) 式中不是考虑 X_n^* 而是考虑变量 $(X_n^* \wedge L)$, 其中 L 是某一常数. 那么, 有

$$\mathbf{E}(X_n^* \wedge L)^p \leqslant q\mathbf{E}[X_n(X_n^* \wedge L)^{p-1}] \leqslant q\|X_n\|_p[\mathbf{E}(X_n^* \wedge L)^p]^{1/q},$$

因此, 由不等式 $\mathbf{E}(X_n^* \wedge L)^p \leqslant L^p < \infty$ 可见

$$\mathbf{E}(X_n^* \wedge L)^p \leqslant q^p \mathbf{E}X_n^p = q^p\|X_n\|_p^p,$$

从而

$$\mathbf{E}(X_n^*)^p = \lim_{L\to\infty} \mathbf{E}(X_n^* \wedge L)^p = q^p\|X_n\|_p^p.$$

现在证明 (10) 式的第二个不等式.

仍然利用 (1) 式, 有

$$\begin{aligned}\mathbf{E}X_n^* - 1 \leqslant \mathbf{E}(X_n^* - 1)^+ &= \int_0^\infty \mathbf{P}\{X_n^* - 1 \geqslant t\}\mathrm{d}t \\ &\leqslant \int_0^\infty \frac{1}{1+t}\left[\int_{\{X_n^* \geqslant 1+t\}} X_n \mathrm{d}\mathbf{P}\right]\mathrm{d}t = \mathbf{E}X_n \int_0^{X_n^*-1} \frac{\mathrm{d}t}{1+t} = \mathbf{E}X_n \cdot \ln X_n^*.\end{aligned}$$

由于对于任何 $a \geqslant 0$ 和 $b > 0$, 有

$$a \ln b \leqslant a \ln^+ a + b\mathrm{e}^{-1}, \tag{19}$$

因此

$$\mathbf{E}X_n^* - 1 \leqslant \mathbf{E}X_n \ln X_n^* \leqslant \mathbf{E}X_n \ln^+ X_n + \mathrm{e}^{-1}\mathbf{E}X_n^*.$$

如果 $\mathbf{E}X_n^* < \infty$, 则由此立即得 (10) 式的第二个不等式.

如果 $\mathbf{E}X_n^* = \infty$, 则像上面一样将 X_n^* 换成 $X_n^* \wedge L$.

证明定理 3 由于 (若 $\mathbf{E}|X|^p < \infty, n \geqslant 0$) $|X|^p (p \geqslant 1)$是非负下鞅, 故由 (1) 式和 (9) 式可见定理 3 成立.

定理 3 的系 *设 $X_n = \xi_0 + \cdots + \xi_n, n \geqslant 0$, 其中 $(\xi_k)_{k\geqslant 0}$ 是独立随机变量序列, 且 $\mathbf{E}\xi_k = 0, \mathbf{E}\xi_k^2 < \infty$. 那么, 不等式 (13) 就是柯尔莫戈洛夫不等式* (第四章 §2).

2. 最大概率及 L^p 中最大范数的估计式 设 $X = (X_n, \mathscr{F}_n)_{n\geqslant 0}$ 是非负下鞅, 而

$$X_n = M_n + A_n$$

是杜布分解 (§1 的 (11) 式). 那么, 由于 $\mathbf{E}M_n = 0$, 故由 (1) 式可见

$$\mathbf{P}\{X_n^* \geqslant \varepsilon\} \leqslant \frac{\mathbf{E}A_n}{\varepsilon}.$$

下面的定理 4 证明此不等式不仅对于下鞅成立, 而且对于控制性的更广泛的一类序列成立. 下面的定义说明了这里所谓*控制性* (*优势性*) 的含义.

定义　设 $X=(X_n,\mathscr{F}_n)$ 是某一非负随机序列, 而 $A=(A_n,\mathscr{F}_{n-1})$ 是一递增可预测序列. 称序列 A 控制序列 X (或 A 是 X 的优势序列), 如果对于任何停时 τ, 有

$$\mathbf{E}X_\tau \leqslant \mathbf{E}A_\tau. \tag{20}$$

定理 4　设 $A=(A_n,\mathscr{F}_{n-1})$ 是一递增可预测序列, 而 $X=(X_n,\mathscr{F}_n)$ 是被序列 A 控制的非负随机序列, 那么, 对于 $\lambda>0, a>0$ 及任何停时 τ, 有.

$$\mathbf{P}\{X_\tau^*\geqslant\lambda\}\leqslant\frac{\mathbf{E}A_\tau}{\lambda}, \tag{21}$$

$$\mathbf{P}\{X_\tau^*\geqslant\lambda\}\leqslant\frac{1}{\lambda}\mathbf{E}(A_\tau\wedge a)+\mathbf{P}\{A_\tau\geqslant a\}, \tag{22}$$

$$\|X_\tau^*\|_p\leqslant\left(\frac{2-p}{1-p}\right)^{1/p}\|A_\tau\|_p,\quad 0<p<1. \tag{23}$$

证明　设

$$\sigma_n=\min\{j\leqslant\tau\wedge n: X_j\geqslant\lambda\},$$

而若 $\{\cdot\}=\varnothing$, 则认为 $\sigma_n=\tau\wedge n$. 那么,

$$\mathbf{E}A_\tau\geqslant\mathbf{E}A_{\sigma_n}\geqslant\mathbf{E}X_{\sigma_n}\geqslant\int_{\{X_{\tau\wedge n}^*>\lambda\}}X_{\sigma_n}\mathrm{d}\mathbf{P}\geqslant\lambda\mathbf{P}\{X_{\tau\wedge n}^*>\lambda\},$$

因而

$$\mathbf{P}\{X_{\tau\wedge n}^*>\lambda\}\leqslant\frac{1}{\lambda}\mathbf{E}A_\tau,$$

于是, 由法图引理可得不等式 (21).

为证明 (22) 式, 引进时间

$$\gamma=\inf\{j: A_{j+1}\geqslant a\},$$

而若 $\{\cdot\}=\varnothing$, 则认为 $\gamma=\infty$. 那么,

$$\begin{aligned}\mathbf{P}\{X_\tau^*\geqslant\lambda\}&=\mathbf{P}\{X_\tau^*\geqslant\lambda, A_\tau<a\}+\mathbf{P}\{X_\tau^*\geqslant\lambda, A_\tau\geqslant a\}\\&\leqslant\mathbf{P}\{I_{\{A_\tau<a\}}X_\tau^*\geqslant\lambda\}+\mathbf{P}\{A_\tau\geqslant a\}\leqslant\mathbf{P}\{X_{\tau\wedge\gamma}^*\geqslant\lambda\}+\mathbf{P}\{A_\tau\geqslant a\}\\&\leqslant\frac{1}{\lambda}\mathbf{E}A_{\tau\wedge\gamma}+\mathbf{P}\{A_\tau\geqslant a\}\leqslant\frac{1}{\lambda}\mathbf{E}(A_\tau\wedge a)+\mathbf{P}\{A_\tau\geqslant a\},\end{aligned}$$

其中用到不等式 (21) 和 $I_{\{A_\tau<a\}}X_\tau^*\leqslant X_{\tau\wedge\gamma}^*$. 最后 (注意到 (22) 式), 由下面的一系列关系式, 得不等式 (23):

$$\begin{aligned}\|X_\tau^*\|_p^p=\mathbf{E}(X_\tau^*)^p&=\int_0^\infty\mathbf{P}\{(X_\tau^*)^p\geqslant t\}\mathrm{d}t=\int_0^\infty\mathbf{P}\{X_\tau^*\geqslant t^{1/p}\}\mathrm{d}t\\&\leqslant\int_0^\infty t^{-1/p}\mathbf{E}[A_\tau\wedge t^{1/p}]\mathrm{d}t+\int_0^\infty\mathbf{P}\{A_\tau^p\geqslant t\}\mathrm{d}t\\&=\mathbf{E}\int_0^{A_\tau^p}\mathrm{d}t+\mathbf{E}\int_{A_\tau^p}^\infty A_\tau t^{-1/p}\mathrm{d}t+\mathbf{E}X_\tau^p=\frac{2-p}{1-p}\mathbf{E}X_\tau^p.\end{aligned}$$

□

注　假设定理 4 的条件成立, 但序列 $A=(A_n,\mathscr{F}_{n\geqslant 0})$ 未必是可预测的, 不过假定对于某个常数 $c>0$, 有

$$\mathbf{P}\left\{\sup_{k\geqslant 1}|\Delta A_k|\leqslant c\right\}=1,$$

其中 $\Delta A_k=A_k-A_{k-1}$. 那么, 如下不等式成立 (对照 (22) 式):

$$\mathbf{P}\{X_\tau^*\geqslant\lambda\}\leqslant\frac{1}{\lambda}\mathbf{E}[A_\tau\wedge(a+c)]+\mathbf{P}\{A_\tau\geqslant a\}. \tag{24}$$

这一事实证明与不等式 (22) 的证明类似, 只需要把时间 $\gamma=\inf\{j:A_{j+1}\geqslant a\}$ 换成时间 $\gamma=\inf\{j:A_j\geqslant a\}$, 并且注意到 $A_\gamma\leqslant a+c$.

系　设 $X^k=(X_n^k,\mathscr{F}_n^k)$ 和 $A^k=(A_n^k,\mathscr{F}_n^k), n\geqslant 0, k\geqslant 1$, 满足定理 4 的条件, 或者满足上面的注. 假设 $(\tau^k)_{k\geqslant 1}$ 关于 $\mathscr{F}^k=(\mathscr{F}_n^k)$ 是停时序列, 而且 $A_{\tau^k}^k\xrightarrow{\mathbf{P}}0$, 那么 $(X^k)_{\tau_k}^*\xrightarrow{\mathbf{P}}0$.

3. 鞅的不等式　在这一小节将 (不加证明, 但是有其应用) 列举一系列非常好的鞅的不等式. 这些不等式, 有些是下面的介绍的辛钦不等式的推广, 有些是独立随机变量之和的马尔钦凯维奇 (J. Matcinkiewicz) 和齐格蒙特 (A. Zygmund) 不等式的推广.

辛钦不等式　设 $\xi_1,\xi_2,\cdots$ 是独立同分布伯努利随机变量, 且 $\mathbf{P}\{\xi_i=1\}=\mathbf{P}\{\xi_i=-1\}=1/2$. 而 $(c_n)_{n\geqslant 1}$ 是一数列.

那么, 对于任意 $0<p<1$ 存在 (不依赖于 (c_n) 的) 通用常数 A_p 和 B_p, 使得对于任意 $n\geqslant 1$, 有

$$A_p\left(\sum_{j=1}^n c_j^2\right)^{1/2}\leqslant\left\|\sum_{j=1}^n c_j\xi_j\right\|_p\leqslant B_p\left(\sum_{j=1}^n c_j^2\right)^{1/2}. \tag{25}$$

马尔钦凯维奇和齐格蒙特不等式　设 $\xi_1,\xi_2,\cdots$ 是独立可积随机变量序列, 且 $\mathbf{E}\xi_i=0$, 则对于任意 $p\geqslant 1$, 存在这样的 (不依赖于 (ξ_n) 的) 通用常数 A_p 和 B_p, 使得对于任意 $n\geqslant 1$, 有

$$A_p\left\|\left(\sum_{j=1}^n \xi_j^2\right)^{1/2}\right\|_p\leqslant\left\|\sum_{j=1}^n \xi_j\right\|_p\leqslant B_p\left\|\left(\sum_{j=1}^n \xi_j^2\right)^{1/2}\right\|_p. \tag{26}$$

在不等式 (25) 和 (26) 中, 序列 $X=(X_n)$ 是鞅, 其中相应地

$$X_n=\sum_{j=1}^n c_j\xi_j\quad 和\quad X_n=\sum_{j=1}^n \xi_j.$$

自然地提出问题, 是否可以将这些不等式推广到任意鞅? 在此方向上最早的结果属于伯克霍尔德 (D. L. Burkholder).

伯克霍尔德不等式 如果 $X=(X_n,\mathscr{F}_n)$ 是鞅, 则对于任意 $p>1$, 存在这样的 (不依赖于 X 的) 通用常数 A_p 和 B_p, 使得对于任意 $n\geqslant 1$, 有

$$A_p\left\|\sqrt{[X]_n}\right\|_p\leqslant\|X_n\|_p\leqslant B_p\left\|\sqrt{[X]_n}\right\|_p,\tag{27}$$

其中 $[X]_n$ 表示 X_n 的二次变差 (第 118 页),

$$[X]_n=\sum_{j=i}^{n}(\Delta X_j)^2,\quad X_0=0.\tag{28}$$

特别, 常数 A_p 和 B_p 可以取为

$$A_p=[18p^{3/2}/(p-1)]^{-1}\quad 和\quad B_p=18p^{3/2}/(p-1)^{1/2}.$$

考虑到 (12) 式, 由 (27) 式可见

$$A_p\left\|\sqrt{[X]_n}\right\|_p\leqslant\|X_n^*\|_p\leqslant B_p^*\left\|\sqrt{[X]_n}\right\|_p,\tag{29}$$

其中

$$A_p=[18p^{3/2}/(p-1)]^{-1}\quad 和\quad B_p^*=18p^{5/2}/(p-1)^{3/2}.$$

当 $p>1$ 时, 伯克霍尔德不等式 (27) 成立; 而且马尔钦凯维奇和齐格蒙特不等式 (26), 也对于 $p=1$ 成立. 问对于 $p=1$, 不等式 (27) 是否成立? 下面的例子表明, 当 $p=1$ 时, 不等式 (27) 不能直接推广.

例 设 $\xi_1,\xi_2,\cdots$ 是独立伯努利随机变量, 而 $\mathbf{P}\{\xi_i=1\}=\mathbf{P}\{\xi_i=-1\}=1/2$, 且

$$X_n=\sum_{j=1}^{n\wedge\tau}\xi_j,\quad 其中\quad \tau=\inf\left\{n\geqslant 1:\sum_{j=1}^{n}\xi_j=1\right\}.$$

序列 $X=(X_n,\mathscr{F}_n)$ 是鞅, 其中

$$\|X_n\|_1=\mathbf{E}|X_n|=2\mathbf{E}X_n^+\to 2,\quad n\to\infty.$$

但是

$$\|\sqrt{[X]_n}\|_1=\mathbf{E}\sqrt{[X]_n}=\mathbf{E}\left(\sum_{j=1}^{n\wedge\tau}1\right)^{1/2}=\mathbf{E}\sqrt{\tau\wedge n}\to\infty.$$

从而, (27) 式的第一个不等式不成立.

结果表明, 对于 $p=1$ 的情形, 虽然不等式 (27) 不能推广, 然而 (对于 $p>1$, 与不等式 (27) 等价的) 不等式 (29) 可以推广.

戴维斯 (H. T. Davis) 不等式 如果 $X=(X_n,\mathscr{F}_n)$ 是鞅, 则存在这样的通用常数 A 和 $B(0<A<B<\infty)$, 使

$$A\left\|\sqrt{[X]_n}\right\|_1\leqslant\|X_n^*\|_1\leqslant B\left\|\sqrt{[X]_n}\right\|_1,\tag{30}$$

即

$$A\mathbf{E}\sqrt{\sum_{j=1}^{n}(\Delta X_j)^2} \leqslant \mathbf{E}\left[\max_{1\leqslant j\leqslant n}|X_j|\right] \leqslant B\mathbf{E}\sqrt{\sum_{j=1}^{n}(\Delta X_j)^2}.$$

系 1 设 $\xi_1,\xi_2,\cdots$ 是独立同分布随机变量, $S_n=\xi_1+\cdots+\xi_n$. 如果 $\mathbf{E}|\xi_1|<\infty, \mathbf{E}\xi_1=0$, 则根据 §2 中瓦尔德不等式 (13), 对于任意 (关于 $(\mathscr{F}_n^\xi)$ 的) 停时 $\tau(\tau<\infty)$, 如下等式成立:

$$\mathbf{E}S_\tau=0. \tag{31}$$

如果补充假设 $\mathbf{E}|\xi_1|^r<\infty(1<r\leqslant 2)$, 则等式 $\mathbf{E}S_\tau=0$ 成立的充分条件是 $\mathbf{E}\tau^{1/r}<\infty$.

为证明系 1 的结论, 引进记号: $\tau_n=\tau\wedge n, Y=\sup\limits_n|S_{\tau_n}|$, 并且假设 $t>0, m=[t^r]$ 是 t^r 的整数部分. 由于 §2 定理 1 的系 1, 可知 $\mathbf{E}S_{\tau_n}=0$. 因此, 为关系式 $\mathbf{E}S_\tau=0$ 成立, (根据控制收敛定理) 只需验证 $\mathbf{E}\sup\limits_n|S_{\tau_n}|<\infty$.

由不等式 (1) 和 (27), 有

$$\begin{aligned}\mathbf{P}\{Y\geqslant t\}&=\mathbf{P}\{\tau\geqslant t^r,Y\geqslant t\}+\mathbf{P}\{\tau<t^r,Y\geqslant t\}\\&\leqslant\mathbf{P}\{\tau\geqslant t^r\}+\mathbf{P}\left\{\max_{1\leqslant j\leqslant m}|S_{\tau_j}|\geqslant t\right\}\leqslant\mathbf{P}\{\tau\geqslant t^r\}+t^{-r}\mathbf{E}|S_{\tau_m}|^r\\&\leqslant\mathbf{P}\{\tau\geqslant t^r\}+t^{-r}B_r^r\mathbf{E}\left(\sum_{j=1}^{\tau_m}\xi_j^2\right)^{r/2}\leqslant\mathbf{P}\{\tau\geqslant t^r\}+t^{-r}B_r^r\mathbf{E}\sum_{j=1}^{\tau_m}|\xi_j|^r.\end{aligned}$$

注意到, (记 $\mathscr{F}_0^\xi=\{\varnothing,\Omega\}$)

$$\begin{aligned}\mathbf{E}\sum_{j=1}^{\tau_m}|\xi_j|^r&=\mathbf{E}\sum_{j=1}^{\infty}I(j\leqslant\tau_m)|\xi_j|^r=\sum_{j=1}^{\infty}\mathbf{E}\mathbf{E}\left[I(j\leqslant\tau_m)|\xi_j|^r\middle|\mathscr{F}_{j-1}^\xi\right]\\&=\mathbf{E}\sum_{j=1}^{\infty}I(j\leqslant\tau_m)\mathbf{E}\left[|\xi_j|^r\middle|\mathscr{F}_{j-1}^\xi\right]=\mathbf{E}\sum_{j=1}^{\tau_m}\mathbf{E}|\xi_j|^r=\mu_r\mathbf{E}\tau_m,\end{aligned}$$

其中 $\mu_r=\mathbf{E}|\xi_1|^r$. 因此,

$$\begin{aligned}\mathbf{P}\{Y\geqslant t\}&\leqslant\mathbf{P}\{\tau\geqslant t^r\}+t^{-r}B_r^r\mu_r\mathbf{E}\tau_m\\&=\mathbf{P}\{\tau\geqslant t^r\}+B_r^r\mu_r t^{-r}\left[m\mathbf{P}\{\tau\geqslant t^r\}+\int_{\{\tau<t^r\}}\tau\mathrm{d}\mathbf{P}\right]\\&\leqslant(1+B_r^r\mu_r)\mathbf{P}\{\tau\geqslant t^r\}+B_r^r\mu_r t^{-r}\int_{\{\tau<t^r\}}\tau\mathrm{d}\mathbf{P},\end{aligned}$$

于是,

$$\mathbf{E}Y = \int_0^\infty \mathbf{P}\{Y \geqslant t\}\mathrm{d}t \leqslant (1 + B_r^r\mu_r)\mathbf{E}\tau^{1/r} + B_r^r\mu_r \int_0^\infty t^{-r}\left[\int_{\{\tau<t^r\}} \tau \mathrm{d}\mathbf{P}\right]\mathrm{d}t$$
$$= (1 + B_r^r\mu_r)\mathbf{E}\tau^{1/r} + B_r^r\mu_r \int_\Omega \tau \left[\int_{\tau^{1/r}}^\infty t^{-r}\mathrm{d}t\right]\mathrm{d}\mathbf{P}$$
$$= \left(1 + B_r^r\mu_r + \frac{B_r^r\mu_r}{r-1}\right)\mathbf{E}\tau^{1/r} < \infty.$$

系 2 设 $M = (M_n)$ 是鞅, 满足条件: 对于某一 $r \geqslant 1, \mathbf{E}|M_n|^{2r} < \infty$, 且

$$\sum_{n=1}^\infty \frac{\mathbf{E}|\Delta M_n|^{2r}}{n^{1+r}} < \infty, \quad (M_0 = 0). \tag{32}$$

那么, 强大数定律成立 (对照第四章 §3 定理 2):

$$\frac{M_n}{n} \to 0 \quad (\mathbf{P} - \text{a.c.}), \quad n \to \infty. \tag{33}$$

当 $r = 1$ 时, (32) 式的证明与第四章 §3 定理 2 的证明方法一样. 具体地说, 设

$$m_n = \sum_{k=1}^n \frac{\Delta M_k}{k}.$$

那么,

$$\frac{M_n}{n} = \frac{\sum_{k=1}^n \Delta M_k}{n} = \frac{1}{n}\sum_{k=1}^n k\Delta m_k,$$

而根据克罗内克引理 (第四章 §3), 级数收敛 $(\mathbf{P} - \text{a.c.})$

$$\frac{1}{n}\sum_{k=1}^n k\Delta m_k \to 0, \quad n \to \infty,$$

的充分条件是, 存在有限极限 $\lim_n m_n (\mathbf{P} - \text{a.c.})$. 而极限 $\lim_n m_n (\mathbf{P} - \text{a.c.})$ 存在, 当且仅当对于任意 $\varepsilon > 0$ (第二章 §10 定理 1 和定理 4), 有

$$\mathbf{P}\left\{\sup_{k\geqslant 1} |m_{n+k} - m_n| \geqslant \varepsilon\right\} \to 0, \quad n \to \infty. \tag{34}$$

由于不等式 (1), 有

$$\mathbf{P}\left\{\sup_{k\geqslant 1} |m_{n+k} - m_n| \geqslant \varepsilon\right\} \leqslant \frac{1}{\varepsilon^2}\sum_{k=n}^\infty \frac{\mathbf{E}(\Delta M_k)^2}{k^2}.$$

从而, 当 $r = 1$ 时, (32) 式和 (34) 式得证.

现在假设 $r>1$. 由第二章 §10 定理 1, 可见关系式 (33) 等价于: 对于任意 $\varepsilon>0$,

$$\varepsilon^{2r}\mathbf{P}\left\{\sup_{j\geqslant n}\frac{|M_j|}{j}\geqslant\varepsilon\right\}\to 0,\quad n\to\infty. \tag{35}$$

由下面 (第 7 小节) 练习题 1 的不等式 (52), 可见

$$\begin{aligned}\varepsilon^{2r}\mathbf{P}\left\{\sup_{j\geqslant n}\frac{|M_j|}{j}\geqslant\varepsilon\right\}&=\varepsilon^{2r}\lim_{m\to\infty}\mathbf{P}\left\{\max_{n\leqslant j\leqslant m}\frac{|M_j|^{2r}}{j^{2r}}\geqslant\varepsilon^{2r}\right\}\\&\leqslant\frac{1}{n^{2r}}\mathbf{E}|M_n|^{2r}+\sum_{j\geqslant n+1}\frac{1}{j^{2r}}\mathbf{E}(|M_j|^{2r}-|M_{j-1}|^{2r}).\end{aligned}$$

由克罗内克引理, 可得

$$\lim_{n\to\infty}\frac{1}{n^{2r}}\mathbf{E}|M_n|^{2r}=0.$$

从而, 为证明 (35) 式只需验证

$$\sum_{j\geqslant 2}\frac{1}{j^{2r}}\mathbf{E}(|M_j|^{2r}-|M_{j-1}|^{2r})<\infty. \tag{36}$$

易见, 有

$$\begin{aligned}I_N&\equiv\sum_{j=2}^{N}\frac{1}{j^{2r}}\mathbf{E}[|M_j|^{2r}-\mathbf{E}|M_{j-1}|^{2r}]\\&\leqslant\sum_{j=2}^{N}\left[\frac{1}{(j-1)^{2r}}-\frac{1}{j^{2r}}\right]\mathbf{E}|M_{j-1}|^{2r}+\frac{\mathbf{E}|M_N|^{2r}}{N^{2r}}.\end{aligned}$$

由伯克霍尔德不等式 (27) 和赫尔德 (O. L. Hölder) 不等式, 可见

$$\mathbf{E}|M_j|^{2r}\leqslant B_{2r}^{2r}\mathbf{E}\left[\sum_{i=1}^{j}(\Delta M_i)^2\right]^r\leqslant B_{2r}^{2r}\mathbf{E}\sum_{i=1}^{j}(\Delta M_i)^{2r}.$$

因而

$$\begin{aligned}I_N&\leqslant\sum_{j=2}^{N-1}B_{2r}^{2r}\left[\frac{1}{j^{2r}}-\frac{1}{(j+1)^{2r}}\right]j^{r-1}\sum_{i=1}^{j}\mathbf{E}|\Delta M_i|^{2r}\frac{\mathbf{E}|M_N|^{2r}}{N^{2r}}\\&\leqslant C_1\sum_{j=2}^{N-1}\frac{1}{j^{r+2}}\sum_{i=1}^{j}\mathbf{E}|\Delta M_i|^{2r}\frac{\mathbf{E}|M_N|^{2r}}{N^{2r}}\leqslant C_2\sum_{j=2}^{N}\frac{\mathbf{E}|\Delta M_j|^{2r}}{j^{r+1}}+C_3.\end{aligned}$$

其中 $C_i(i=1,2,3)$ 是常数. 于是, 由于 (32) 式, 估计式 (36) 得证.

4. 下鞅的极限平均 "振动" 次数的上界 随机变量序列 $(X_n)_{n\geqslant 1}$ 以概率 1 有 (有限或无限) 极限 $\lim X_n$, 当且仅当 "在任意两个 (有理) 数 a 和 $b(a<b)$ 之间振动的" 次数以概率 1 有限. 对于下鞅, 下面的定理 5 给出了 "振动的" 平均次数这上侧估计, 定理 5 将用来证明其收敛性的基本结果.

固定两个 (有理) 数 a 和 $b(a<b)$, 并且对于序列 $X=(X_n)_{n\geqslant 1}$ 定义停时:

$$\begin{aligned}\tau_0&=0,\\ \tau_1&=\min\{n>0:X_n\leqslant a\},\\ \tau_2&=\min\{n>\tau_1:X_n\geqslant b\},\\ &\cdots\cdots\cdots\\ \tau_{2m-1}&=\min\{n>\tau_{2m-2}:X_n\leqslant a\},\\ \tau_{2m}&=\min\{n>\tau_{2m-1}:X_n\geqslant b\},\end{aligned}$$

若相应的集合 $\{\cdot\}$ 是空集, 则设 $\tau_k=\infty$.

其次, 对于每一个 $n\geqslant 1$, 定义随机变量

$$\beta_n(a,b)=\begin{cases}0, & \text{若 } \tau_2>n,\\ \max\{m:\tau_{2m}\leqslant n\}, & \text{若 } \tau_2\leqslant n.\end{cases}$$

随机变量 $\beta_n(a,b)$ 的含义是: 序列 $X_1,\cdots,X_n$ 与区间 $[a,b]$ (自下而上) 相交的次数.

定理 5 (杜布) 设 $X=(X_n,\mathscr{F}_n)_{n\geqslant 1}$ 是下鞅. 那么, 对于任意 $n\geqslant 1$, 有

$$\mathbf{E}\beta_n(a,b)\leqslant\frac{\mathbf{E}|X_n-a|^+}{b-a}. \tag{37}$$

证明 下鞅 $X=(X_n,\mathscr{F}_n)$ 与区间 $[a,b]$ 相交的次数, 等于非负下鞅 $X^+=((X_n-a)^+,\mathscr{F}_n)$ 与区间 $[0,b-a]$ 相交的次数. 因此, 如果把下鞅 X 视为非负的且 $a=0$, 则需要证明

$$\mathbf{E}\beta_n(0,b)\leqslant\frac{\mathbf{E}X_n}{b}. \tag{38}$$

设 $X_0=0,F_0=\{\varnothing,\Omega\}$, 而对于 $i=1,2,\cdots$ 和某个数 m, 记

$$\varphi_i=\begin{cases}1, & \text{若 } \tau_m<i\leqslant\tau_{m+1} \text{ 且 } m \text{ 为奇数},\\ 0, & \text{若 } \tau_m<i\leqslant\tau_{m+1} \text{ 且 } m \text{ 为偶数}.\end{cases}$$

不难看到,

$$b\beta_n(0,b)\leqslant\sum_{i=1}^n\varphi_i[X_i-X_{i-1}]$$

和

$$\{\varphi_i=1\}=\bigcup_{m\ \text{为奇数}}[\{\tau_m<i\}\setminus\{\tau_{m+1}<i\}]\in\mathscr{F}_{i-1}.$$

从而

$$
\begin{aligned}
b\mathbf{E}\beta_n(0,b) &\leqslant \mathbf{E}\sum_{i=1}^{n}\varphi_i[X_i - X_{i-1}] = \sum_{i=1}^{n}\int_{\{\varphi_i=1\}}(X_i - X_{i-1})\mathrm{d}\mathbf{P} \\
&= \sum_{i=1}^{n}\int_{\{\varphi_i=1\}}\mathbf{E}(X_i - X_{i-1}|\mathscr{F}_{i-1})\mathrm{d}\mathbf{P} = \sum_{i=1}^{n}\int_{\{\varphi_i=1\}}[\mathbf{E}(X_i|\mathscr{F}_{i-1}) - X_{i-1}]\mathrm{d}\mathbf{P} \\
&\leqslant \sum_{i=1}^{n}\int_{\Omega}[\mathbf{E}(X_i|\mathscr{F}_{i-1}) - X_{i-1}]\mathrm{d}\mathbf{P} = \mathbf{E}X_n,
\end{aligned}
$$

于是, 不等式 (38) 得证. □

5. 二次可积鞅大偏差概率的估计 在这一小节对于平方可积鞅, 我们将讨论大偏差概率的不等式的某些简单不等式.

设 $M=(M_n,\mathscr{F}_n)_{n\geqslant 0}$ 是平方可积鞅, 而 $\langle M\rangle = (\langle M\rangle_n,\mathscr{F}_{n-1}), M_0=0$ 是其二次特征. 如果将不等式 (22) 用于 $X_n = M_n^2, A_n = \langle M\rangle_n$, 则对于 $a>0, b>0$, 得

$$
\begin{aligned}
&\mathbf{P}\left\{\max_{k\leqslant n}|M_k| \geqslant an\right\} = \mathbf{P}\left\{\max_{k\leqslant n}M_k^2 \geqslant (an)^2\right\} \\
&\leqslant \frac{1}{(an)^2}\mathbf{E}[\langle M\rangle_n \wedge (bn)] + \mathbf{P}\{\langle M\rangle_n \geqslant an\}. \qquad (39)
\end{aligned}
$$

事实上, 利用第四章 §5, 在估计独立同分布随机变量之和的、大偏差概率的估计时, 所阐述的思想, 至少当 $|\Delta M_n| \leqslant C$ 时, 对于一切 n 和 $\omega\in\Omega$, 可以本质上改进该不等式.

注意, 在第四章 §5 中推导相应的不等式时, 这样的环节是利用了如下事实: 序列

$$
(\mathrm{e}^{\lambda S_n}/[\varphi(\lambda)]^n, \mathscr{F}_n)_{n\geqslant 1}, \quad \mathscr{F}_n = \sigma(\xi_1,\cdots,\xi_n), \qquad (40)
$$

形成非负鞅, 然后再对其运用本节的不等式 (8). 如果现在将 S_n 取作 M_n, 则与 (40) 式类似的将是非负鞅

$$
(\mathrm{e}^{\lambda M_n}/\mathscr{E}_n(\lambda), \mathscr{F}_n)_{n\geqslant 1},
$$

其中

$$
\mathscr{E}_n(\lambda) = \prod_{j=1}^{n}\mathbf{E}\left(\mathrm{e}^{\lambda\Delta M_j}|\mathscr{F}_{j-1}\right) \qquad (41)
$$

是所谓随机分量 (亦见第二章 §6 第 13 小节).

这一表达式相当复杂. 然而, 为使所形成的序列是鞅, 完全没有必要利用不等式 (8). 只需使它形成非负上鞅. 我们在这里也正是这样做的: 首先构造序列 $(Z_n(\lambda),\mathscr{F}_n)$ (见下面的 (43) 式), 然后运用第四章 §5 使用的方法.

引理 1　设 $M=(M_n,\mathscr{F}_n)_{n\geqslant 0}$ 是二次可积鞅, $M_0=0,\Delta M_0=0$, 而对于一切 n 和 $\omega,|\Delta M_n(\omega)|\leqslant c$. 假设对于 $\lambda>0$, 有

$$\psi_c(\lambda)=\begin{cases}\dfrac{\mathrm{e}^{\lambda c}-1-\lambda c}{c^2}, & c>0,\\ \dfrac{\lambda^2}{2}, & c\leqslant 0,\end{cases}\tag{42}$$

而

$$Z_n(\lambda)=\mathrm{e}^{\lambda M_n-\psi_c(\lambda)\langle M\rangle_n}.\tag{43}$$

那么, 对于每一个 $c\geqslant 0$, 序列 $Z(\lambda)=(Z_n(\lambda),\mathscr{F}_n)_{n\geqslant 0}$ 是非负上鞅.

证明　对于 $|x|\leqslant c$,

$$\mathrm{e}^{\lambda x}-1-\lambda x=(\lambda x)^2\sum_{m\geqslant 2}\frac{(\lambda x)^{m-2}}{m!}\leqslant(\lambda x)^2\sum_{m\geqslant 2}\frac{(\lambda c)^{m-2}}{m!}\leqslant x^2\psi_c(\lambda).$$

由于该不等式以及表达式 $(Z_n=Z_n(\lambda))$

$$\Delta Z_n=Z_{n-1}[(\mathrm{e}^{\lambda\Delta M_n}-1)\mathrm{e}^{-\Delta\langle M\rangle_n\psi_c(\lambda)}+(\mathrm{e}^{-\Delta\langle M\rangle_n\psi_c(\lambda)}-1)],$$

得

$$\begin{aligned}&\mathbf{E}(\Delta Z_n|\mathscr{F}_{n-1})\\ &=Z_{n-1}[\mathbf{E}(\mathrm{e}^{\lambda\Delta M_n}-1|\mathscr{F}_{n-1})\mathrm{e}^{-\Delta\langle M\rangle_n\psi_c(\lambda)}+(\mathrm{e}^{-\Delta\langle M\rangle_n\psi_c(\lambda)}-1)]\\ &=Z_{n-1}[\mathbf{E}(\mathrm{e}^{\lambda\Delta M_n}-1-\lambda\Delta M_n|\mathscr{F}_{n-1})\mathrm{e}^{-\Delta\langle M\rangle_n\psi_c(\lambda)}+(\mathrm{e}^{-\Delta\langle M\rangle_n\psi_c(\lambda)}-1)]\\ &\leqslant Z_{n-1}[\psi_c(\lambda)\mathbf{E}((\Delta M_n)^2|\mathscr{F}_{n-1})\mathrm{e}^{-\Delta\langle M\rangle_n\psi_c(\lambda)}+(\mathrm{e}^{-\Delta\langle M\rangle_n\psi_c(\lambda)}-1)]\\ &=Z_{n-1}[\psi_c(\lambda)\Delta\langle M\rangle_n\mathrm{e}^{-\Delta\langle M\rangle_n\psi_c(\lambda)}+(\mathrm{e}^{-\Delta\langle M\rangle_n\psi_c(\lambda)}-1)]\leqslant 0,\end{aligned}\tag{44}$$

其中亦用到如下不等式: 对于 $x\geqslant 0$, 有

$$x\mathrm{e}^{-x}+(\mathrm{e}^{-x}-1)\leqslant 0.$$

由 (44) 式可见

$$\mathbf{E}(Z_n|\mathscr{F}_{n-1})\leqslant Z_{n-1},$$

即 $Z(\lambda)=(Z_n(\lambda),\mathscr{F}_n)$ 是上鞅.　□

假设满足引理 1 的条件. 那么, 总存在 $\lambda>0$, 使 (对于给定的 $a>0,b>0$), 有

$a\lambda - b\psi_c(\lambda) > 0$. 由此, 可见

$$\begin{aligned}
&\mathbf{P}\left\{\max_{k\leqslant n} M_k \geqslant an\right\} = \mathbf{P}\left\{\max_{k\leqslant n} \mathrm{e}^{\lambda M_k} \geqslant \mathrm{e}^{\lambda an}\right\}\\
&\leqslant \mathbf{P}\left\{\max_{k\leqslant n} \mathrm{e}^{\lambda M_k - \psi_c(\lambda)\langle M\rangle_k} \geqslant \mathrm{e}^{\lambda an - \psi_c(\lambda)\langle M\rangle_n}\right\}\\
&= \mathbf{P}\left\{\max_{k\leqslant n} \mathrm{e}^{\lambda M_k - \psi_c(\lambda)\langle M\rangle_k} \geqslant \mathrm{e}^{\lambda an - \psi_c(\lambda)\langle M\rangle_n}, \langle M\rangle_n \leqslant bn\right\}\\
&\quad + \mathbf{P}\left\{\max_{k\leqslant n} \mathrm{e}^{\lambda M_k - \psi_c(\lambda)\langle M\rangle_k} \geqslant \mathrm{e}^{\lambda an - \psi_c(\lambda)\langle M\rangle_n}, \langle M\rangle_n > nb\right\}\\
&\leqslant \mathbf{P}\left\{\max_{k\leqslant n} \mathrm{e}^{\lambda M_k - \psi_c(\lambda)\langle M\rangle_k} \geqslant \mathrm{e}^{\lambda an - \psi_c(\lambda)bn}\right\} + \mathbf{P}\left\{\langle M\rangle_n > bn\right\}\\
&\leqslant \mathrm{e}^{-n[\lambda a - b\psi_c(\lambda)]} + \mathbf{P}\{\langle M\rangle_n > bn\},
\end{aligned} \tag{45}$$

其中由 (7) 式得最后一个不等式.

记 (对照第四章 §5 的函数 $H(a)$)

$$H_c(a,b) = \sup_{\lambda>0}[\lambda a - b\psi_c(\lambda)].$$

那么, 由 (45) 式可见

$$\mathbf{P}\left\{\max_{k\leqslant n} M_k \geqslant an\right\} \leqslant \mathbf{P}\left\{\langle M\rangle_n > bn\right\} + \mathrm{e}^{-nH_c(a,b)}. \tag{46}$$

将鞅 M 换成 $-M$ 可见, 不等式 (46) 的右侧也从上侧估计概率 $\mathbf{P}\left\{\min\limits_{k\leqslant n} M_k \leqslant -an\right\}$. 因此,

$$\mathbf{P}\left\{\max_{k\leqslant n} |M_k| \geqslant an\right\} \leqslant 2\mathbf{P}\left\{\langle M\rangle_n > bn\right\} + 2\mathrm{e}^{-nH_c(a,b)}. \tag{47}$$

于是, 证明了如下定理.

定理 6 *设 $M = (M_n, \mathscr{F}_n)$ 是具有一致有界跃度的鞅, 即对于某个常数 $c > 0$ 以及一切 n 和 ω, $|\Delta M_n(\omega)| \leqslant c$. 那么, 对于任意 $a > 0, b > 0$, 不等式 (46) 和 (47) 成立.*

注 函数

$$H_c(a,b) = \frac{1}{c}\left(a + \frac{b}{c}\right)\ln\left(1 + \frac{ac}{b}\right) - \frac{a}{c}. \tag{48}$$

6. 二次可积鞅与其二次特征最大比值概率的估计 在定理 6 的条件下, 我们现在考虑形如

$$\mathbf{P}\left\{\sup_{k\geqslant n} \frac{M_k}{\langle M\rangle_k} > a\right\}$$

的概率的估计问题. 特别, 这种形式的概率, 在关于鞅的强大数定律中 (下面见 §5 的定理 4), 表征收敛的速度.

由与第四章 §5 同样的方法, 可见对于任意 $a>0$, 存在 $\lambda>0$, 使 $a\lambda-\psi_c(\lambda)>0$. 那么, 对于任意 $b>0$, 有

$$\begin{aligned}\mathbf{P}\left\{\sup_{k\geqslant n}\frac{M_k}{\langle M\rangle_k}>a\right\}&\leqslant\mathbf{P}\left\{\sup_{k\geqslant n}\mathrm{e}^{\lambda M_k-\psi_c(\lambda)\langle M\rangle_k}>\mathrm{e}^{[a\lambda-\psi_c(\lambda)]\langle M\rangle_n}\right\}\\&\leqslant\mathbf{P}\left\{\sup_{k\geqslant n}\mathrm{e}^{\lambda M_k-\psi_c(\lambda)\langle M\rangle_k}>\mathrm{e}^{[a\lambda-\psi_c(\lambda)]bn}\right\}+\mathbf{P}\left\{\langle M\rangle_n<bn\right\}\\&\leqslant\mathrm{e}^{-bn[a\lambda-\psi_c(\lambda)]}+\mathbf{P}\{\langle M\rangle_n<bn\}.\end{aligned}\tag{49}$$

由此, 得

$$\mathbf{P}\left\{\sup_{k\geqslant n}\frac{M_k}{\langle M\rangle_k}>a\right\}\leqslant\mathbf{P}\left\{\langle M\rangle_n<bn\right\}+\mathrm{e}^{-nH_c(ab,b)},\tag{50}$$

$$\mathbf{P}\left\{\sup_{k\geqslant n}\left|\frac{M_k}{\langle M\rangle_k}\right|>a\right\}\leqslant2\mathbf{P}\left\{\langle M\rangle_n<bn\right\}+2\mathrm{e}^{-nH_c(ab,b)}.\tag{51}$$

于是, 证明了下面的定理.

定理 7　*假设满足定理 6 的条件, 则对于任意 $a>0,b>0$, 不等式 (50) 和 (51) 成立.*

注　将 (51) 式的估计, 与第四章 §5 中的 (21) 式比较, 其中对于伯努利概型: $p=1/2,M_n=S_n-n/2,b=1/4,c=1/2$, 则可以看到当 $\varepsilon>0$ 二者产生同样的结果:

$$\mathbf{P}\left\{\sup_{k\geqslant n}\left|\frac{M_k}{\langle M\rangle_k}\right|>\varepsilon\right\}=\mathbf{P}\left\{\sup_{k\geqslant n}\left|\frac{S_k-k/2}{k}\right|>\frac{\varepsilon}{4}\right\}\leqslant2\mathrm{e}^{-4\varepsilon^2n}.$$

7. 练习题

1. 设 $X=(X_n,\mathscr{F}_n)$ 是非负下鞅, $V=(V_n,\mathscr{F}_{n-1})$ 是可预测序列, 其中以概率 1, 有 $0\leqslant V_{n+1}\leqslant V_n\leqslant C$, 而 C 是常数, 证明不等式 (1) 有如下推广:

$$\varepsilon\mathbf{P}\left\{\max_{1\leqslant j\leqslant n}V_jX_j\geqslant\varepsilon\right\}+\int_{\left\{\max\limits_{1\leqslant j\leqslant n}V_jX_j<\varepsilon\right\}}V_nX_n\mathrm{dP}\leqslant\sum_{j=1}^n\mathbf{E}V_j\Delta X_j.\tag{52}$$

2. 证明*克里克伯格* (Krickbeg) *分解*: 任何鞅 $X=(X_n,\mathscr{F}_n)$, 只要 $\sup\mathbf{E}|X_n|<\infty$, 都可以表示为两个非负鞅之差.

3. 设 $\xi_1,\xi_2,\cdots$ 是独立随机变量序列, $S_n=\xi_1+\cdots+\xi_n$, 而 $S_{m,n}=\sum\limits_{j=m+1}^n\xi_j$. 证明如下*奥塔维安尼* (Ottawiani) *不等式*:

$$\mathbf{P}\left\{\max_{1\leqslant j\leqslant n}|S_j|>2\varepsilon\right\}\leqslant\frac{\mathbf{P}\{|S_n|>\varepsilon\}}{\max\limits_{1\leqslant j\leqslant n}\mathbf{P}\{|S_{j,n}|\leqslant\varepsilon\}},$$

并 (在 $\mathbf{E}\xi_i=0,i\geqslant1$, 的条件下) 导出不等式:

$$\int_0^\infty\mathbf{P}\left\{\max_{1\leqslant j\leqslant n}|S_j|>2t\right\}\mathrm{d}t\leqslant2\mathbf{E}|S_n|+2\int_{2\mathbf{E}|S_n|}^\infty\mathbf{P}\{|S_n|>t\}\mathrm{d}t.\tag{53}$$

4. 设 $\xi_1,\xi_2,\cdots$ 是独立随机变量序列, 且 $\mathbf{E}\xi_i=0$. 利用 (53) 式证明, 在这种情形下有不等式 (10) 的加强:

$$\mathbf{E}S_n^*\leqslant 8\mathbf{E}|S_n|.$$

5. 证明 (16) 式.

6. 证明 (19) 式.

7. 设 $\sigma-$ 代数 $\mathscr{F}_0,\cdots,\mathscr{F}_n$ 满足 $\mathscr{F}_0\subseteq\mathscr{F}_1\subseteq\cdots\subseteq\mathscr{F}_n$, 而事件 $A_k\in\mathscr{F}_k(k=1,\cdots,n)$. 利用 (22) 式证明如下德沃列茨基 (Дворецкий) 不等式: 对于任意 $\varepsilon>0$,

$$\mathbf{P}\left(\bigcup_{k=1}^n A_k\right)\leqslant\varepsilon+\mathbf{P}\left\{\sum_{k=1}^n\mathbf{P}(A_k|\mathscr{F}_{k-1})>\varepsilon\right\}.$$

8. 设 $X=(X_n)_{n\geqslant1}$ 是二次可积鞅, 而 $(b_n)_{n\geqslant1}$ 是正实数不减序列. 证明如下哈伊克－雷内伊 (J. Hajek– A. Rènyi) 不等式: 对于任意 $\lambda>0$,

$$\mathbf{P}\left\{\max_{1\leqslant j\leqslant n}\left|\frac{X_k}{b_k}\right|\geqslant\lambda\right\}\leqslant\frac{1}{\lambda^2}\sum_{k=1}^n\frac{\mathbf{E}(\Delta X_k)^2}{b_k^2},\quad \Delta X_k=X_k-X_{k-1},\quad X_0=0.$$

9. 设 $X=(X_n)_{n\geqslant1}$ 是下鞅, 而 $g(x)$ 是非负递增凹 (下凸) 函数. 那么, 对于任意正 t, 和实数 x, 有

$$\mathbf{P}\left\{\max_{1\leqslant j\leqslant n}X_k\geqslant x\right\}\leqslant\frac{\mathbf{E}g(tX_n)}{g(tx)}.$$

特别,

$$\mathbf{P}\left\{\max_{1\leqslant j\leqslant n}X_k\geqslant x\right\}\leqslant \mathrm{e}^{-tx}\mathbf{E}\mathrm{e}^{tX_n}.$$

10. 设 $\xi_1,\xi_2,\cdots$ 是独立随机变量序列, 且 $\mathbf{E}\xi_m=0,\mathbf{E}\xi_n^2=1,n\geqslant1$. 记

$$\tau=\inf\left\{n\geqslant1:\sum_{i=1}^n\xi_i>0\right\}.$$

证明 $\mathbf{E}\tau^{1/2}<\infty$.

11. 设 $\xi=(\xi_n)_{n\geqslant1}$ 是鞅－差, 而 $1<p\leqslant2$. 证明

$$\mathbf{E}\sup_{n\geqslant1}\left|\sum_{j=1}^n\xi_j\right|^p\leqslant C_p\sum_{j=1}^n\mathbf{E}|\xi_j|^p,$$

其中 C_p是常数.

12. 设 $X=(X_n)_{n\geqslant1}$ 是鞅, $\mathbf{E}\xi_m=0,\mathbf{E}\xi_n^2<\infty,n\geqslant1$. 证明 (第四章 §2 练习题 5), 对于任意 $n\geqslant1$ 和 $\varepsilon>0$, 有

$$\mathbf{P}\left\{\max_{1\leqslant k\leqslant n}X_k\geqslant\varepsilon\right\}\leqslant\frac{\mathbf{E}X_n^2}{\varepsilon^2+\mathbf{E}X_n^2}.$$

§4. 下鞅和鞅收敛的基本定理

1. 有界单调鞅序列极限的存在性 下面的结果在整个下鞅收敛性问题中, 可以视为数学分析中著名事实 "有界单调数列有 (有限) 极限" 的概率类似,

定理 1 (杜布) 设 $X=(X_n,\mathscr{F}_n)_{n\geqslant 1}$ 是下鞅, 且

$$\sup_n \mathbf{E}|X_n|<\infty, \tag{1}$$

则以概率 1 存在极限 $\lim X_n=X_\infty$, 并且 $\mathbf{E}|X_\infty|<\infty$.

证明 假设

$$\mathbf{P}\left\{\overline{\lim}X_n>\underline{\lim}X_n\right\}>0. \tag{2}$$

那么, 由于

$$\left\{\overline{\lim}X_n>\underline{\lim}X_n\right\}=\bigcup_{a<b}\left\{\overline{\lim}X_n>b>a>\underline{\lim}X_n\right\}$$

(a,b 是有理数), 故存在 a 和 b, 使

$$\mathbf{P}\left\{\overline{\lim}X_n>b>a>\underline{\lim}X_n\right\}>0. \tag{3}$$

设 $\beta_n(a,b)$ 是序列 $X_1,\cdots,X_n$ 自下而上与区间 (a,b) 相交的次数, 而 $\beta_\infty(a,b)=\lim\limits_n\beta_n(a,b)$. 根据 §3 的 (37) 式, 有

$$\mathbf{E}\beta_n(a,b)=\frac{\mathbf{E}|X_n-a|^+}{b-a}\leqslant\frac{\mathbf{E}X_n^++|a|}{b-a},$$

因此,

$$\mathbf{E}\beta_\infty(a,b)=\lim_n\mathbf{E}\beta_n(a,b)\leqslant\frac{\sup\limits_n\mathbf{E}X_n^++|a|}{b-a}<\infty.$$

由 (1) 式以及 (由于 $\mathbf{E}X_n^+\leqslant\mathbf{E}|X_n|=2\mathbf{E}X_n^+-\mathbf{E}X_n\leqslant 2\mathbf{E}X_n^+-\mathbf{E}X_1$) 对于下鞅, 有

$$\sup_n\mathbf{E}|X_n|<\infty\Leftrightarrow\sup_n\mathbf{E}X_n^+<\infty.$$

然而, 条件 $\mathbf{E}\beta_\infty(a,b)<\infty$ 与假设 (3) 矛盾. 于是, 以概率 1 存在极限 $\lim X_n=X_\infty$, 而由法图引理可见

$$\mathbf{E}|X_\infty|\leqslant\sup_n\mathbf{E}|X_n|<\infty.$$

□

系 1 如果 X 是非负下鞅, 则以概率 1 存在有限极限 $\lim X_n=X_\infty$.

系 2 如果 $X=(X_n,\mathscr{F}_n)_{n\geqslant 1}$ 是非正下鞅, 则序列

$$\overline{X}=(X_n,\mathscr{F}_n),\quad 1\leqslant n\leqslant\infty,\quad X_\infty=\lim X_n \quad 和 \quad \mathscr{F}_\infty=\sigma\left(\bigcup_n\mathscr{F}_n\right)$$

构成 (非正) 下鞅.

事实上, 根据法图引理,

$$\mathbf{E}X_\infty = \mathbf{E}\lim X_n \geqslant \overline{\lim}\mathbf{E}X_n > \mathbf{E}X_1 > -\infty,$$

且以概率 1, 有

$$\mathbf{E}(X_\infty|\mathscr{F}_m) = \mathbf{E}(\lim X_n|\mathscr{F}_m) \geqslant \overline{\lim}\mathbf{E}(X_n|\mathscr{F}_m) \geqslant X_m.$$

系 3 如果 $X = (X_n, \mathscr{F}_n)$ 是非负上鞅, 则以概率 1 存在有限极限 $\lim X_n$.

事实上, 这时有

$$\sup_n \mathbf{E}|X_n| = \sup_n \mathbf{E}X_n = \mathbf{E}X_1 < \infty,$$

从而, 可以运用定理 1.

2. 鞅几乎必然收敛也是在 L^1 上平均收敛的条件 设 $\xi_1, \xi_2, \cdots$ 是独立随机变量序列, 且 $\mathbf{P}\{\xi_i = 0\} = \mathbf{P}\{\xi_i = 2\} = 1/2$. 那么, $X = (X_n, \mathscr{F}_n^\xi)$ 是鞅, 其中

$$X_n = \prod_{i=1}^n \xi_i, \quad \mathscr{F}_n^\xi = \sigma(\xi_1, \cdots, \xi_n),$$

且 $\mathbf{E}X_n = 1$ 和 $X_n \to X_\infty \equiv 0$ $(\mathbf{P}-\text{a.c.})$. 同样, 显然 $\mathbf{E}|X_n - \mathbf{E}X_\infty| = 1$, 从而 $X_n \nrightarrow X_\infty$. 这样, 条件 (1) 一般不能保障在 L^1 的意义上 X_n 收敛于 X_∞.

下面将要介绍的定理 2 表明, 如果将条件 (1) 加强为 "随机变量族 $\{X_n\}$ 具有一致可积性" (那么, 根据第二章 §6 的第 5 小节知, 性质 (16) 成立), 则 $\{X_n\}$ 在 L_1 意义上几乎处处收敛, 同时也在 L^1 的意义上平均收敛.

定理 2 设 $X = (X_n, \mathscr{F}_n)$ 是下鞅, 且随机变量族 $\{X_n\}$ 一致可积. 那么, 存在随机变量 X_∞, 且 $\mathbf{E}|X_\infty| < \infty$, 使当 $n \to \infty$ 时, 有

$$X_n \to X_\infty \ (\mathbf{P}-\text{a.c.}), \tag{4}$$

$$X_n \xrightarrow{L^1} X_\infty. \tag{5}$$

这时, 序列 $\overline{X} = (X_n, \mathscr{F}_n), 1 \leqslant n \leqslant \infty, \mathscr{F}_\infty = \sigma\left(\bigcup_n \mathscr{F}_n\right)$ 也构成下鞅.

证明 由定理 1 得 (4) 式, 由 (4) 式和第二章 §6 的定理 4 的 (5) 式.

其次, 如果 $A \in \mathscr{F}_n$ 和 $m \geqslant n$, 则

$$\mathbf{E}I_A|X_m - X_\infty| \to 0, \quad m \to \infty,$$

从而,

$$\lim_{n\to\infty} \int_A X_m \mathrm{d}\mathbf{P} = \int_A X_\infty \mathrm{d}\mathbf{P}.$$

由于序列

$$\left(\int_A X_m \mathrm{dP}\right)_{m\geqslant n}$$

是非降的, 可见

$$\int_A X_n \mathrm{dP} \leqslant \int_A X_m \mathrm{dP} \leqslant \int_A X_\infty \mathrm{dP},$$

于是, 对于一切 $n \geqslant 1, X_n \leqslant (X_\infty|\mathscr{F}_n)(\mathbf{P}-\text{a.c.})$. □

系 设 $X = (X_n, \mathscr{F}_n)$ 是下鞅, 且对于某个 $p > 1$,

$$\sup_n \mathbf{E}|X_n| < \infty, \tag{6}$$

则存在可积随机变量 X_∞, 使 (4) 式和 (5) 式成立.

对于证明只需注意到, 根据第二章 §6 的引理 3, 条件 (6) 可以保障随机变量族 $\{X_n\}$ 的一致可积性.

3. 鞅一致可积的充分和必要条件 现在引进关于条件数学期望连续性的定理, 是关于鞅的收敛性的最早的成果之一.

定理 3 (P. 列维 (P. P. Lévy)) 设 $(\Omega, \mathscr{F}, \mathbf{P})$ 是概率空间, $(\mathscr{F}_n)_{n\geqslant 1}$ 是非降 $\sigma-$ 代数族: $\mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \cdots \subseteq \mathscr{F}$. 假设 ξ 是随机变量, 且 $\mathbf{E}|\xi| < \infty$, 而 $\mathscr{F}_\infty = \sigma\left(\bigcup_n \mathscr{F}_n\right)$. 那么, 以概率 1 并且在 L^1 收敛意义上, 有

$$\mathbf{E}(\xi|\mathscr{F}_n) \to \mathbf{E}(\xi|\mathscr{F}_\infty), \quad n \to \infty. \tag{7}$$

证明 设 $X_n = (\xi, \mathscr{F}_n), n \geqslant 1$. 那么, 对于 $a > 0, b > 0$, 有

$$\begin{aligned}
\int_{\{|X_n|\geqslant a\}} |X_n| \mathrm{dP} &\leqslant \int_{\{|X_n|\geqslant a\}} \mathbf{E}(|\xi|\big|\mathscr{F}_n) \mathrm{dP} = \int_{\{|X_n|\geqslant a\}} |\xi| \mathrm{dP} \\
&= \int_{\{|X_n|\geqslant a, |\xi|\leqslant b\}} |\xi| \mathrm{dP} + \int_{\{|X_n|\geqslant a, |\xi|>b\}} |\xi| \mathrm{dP} \\
&\leqslant b\mathbf{P}\{|X_n| \geqslant a\} + \int_{\{|\xi|>b\}} |\xi| \mathrm{dP} \\
&\leqslant \frac{b}{a}\mathbf{E}|\xi| + \int_{\{|\xi|>b\}} |\xi| \mathrm{dP}.
\end{aligned}$$

先令 $a \to \infty$, 再令 $b \to \infty$, 得

$$\lim_{a\to\infty} \sup_n \int_{\{|X_n|\geqslant a\}} |X_n| \mathrm{dP} = 0.$$

这说明随机变量族 $\{X_n\}$ 一致可积. 那么, 根据定理 2, 存在随机变量 X_∞, 使以概率 1 收敛并且在 L^1 中的收敛, 有 $X_n = \mathbf{E}(\xi|\mathscr{F}_n) \to X_\infty$. 因此, 只需设

$$X_\infty = \mathbf{E}(\xi|\mathscr{F}_\infty) \ \ (\mathbf{P}-\text{a.c.}).$$

设 $m \geqslant n, A \in \mathscr{F}_n$, 则

$$\int_A X_m \mathrm{d}\mathbf{P} = \int_A X_n \mathrm{d}\mathbf{P} = \int_A \mathbf{E}(\xi|\mathscr{F}_n)\mathrm{d}\mathbf{P} = \int_A \xi \mathrm{d}\mathbf{P}.$$

由于随机变量族 $\{X_n\}$ 的一致可积, 且由第二章 §6 定理 5 知: 当 $m \to \infty$ 时, 有 $\mathbf{E}I_{\mathbf{A}}|X_m - X_\infty| \to 0$, 从而

$$\int_A X_\infty \mathrm{d}\mathbf{P} = \int_A \xi \mathrm{d}\mathbf{P}. \tag{8}$$

该等式对于任意 $A \in \mathscr{F}_n$ 成立, 因而对于任意 $A \in \bigcup_{n=1}^{\infty} \mathscr{F}_n$ 成立. 由于 $\mathbf{E}|X_\infty| < \infty, \mathbf{E}|\xi| < \infty$, 则 (8) 式的左右两侧都是亦可能取负值的 $\sigma-$ 可加测度, 不过这样的测度是有限的, 并且在代数 $\bigcup_{n=1}^{\infty} \mathscr{F}$ 上重合. 根据卡拉泰奥多里 (C. Carathéodory) 定理 (第二章 §3), 由于测度自代数到 $\sigma-$ 代数上的延拓的唯一性, 可见对于集合 $A \in F_\infty = \sigma(\bigcup F_n)$ 下面的等式 (9) 成立:

$$\int_A X_\infty \mathrm{d}\mathbf{P} = \int_A \xi \mathrm{d}\mathbf{P} = \int_A \mathbf{E}(\xi|\mathscr{F}_\infty)\mathrm{d}\mathbf{P}, \quad A \in \mathscr{F}_\infty. \tag{9}$$

随机变量 X_∞ 和 $\mathbf{E}(\xi|\mathscr{F}_\infty)$ 是 $\mathscr{F}_\infty-$ 可测的, 因此由 (9) 式根据第二章 §6 第 3 小节的性质 I, 可见 $X_\infty = \mathbf{E}(\xi|\mathscr{F}_\infty)(\mathbf{P} - \text{a.c.})$. □

系 *随机序列 $X = (X_n, \mathscr{F}_n)$ 是一致可积鞅, 当且仅当存在随机变量 $\xi, \mathbf{E}|\xi| < \infty$, 使对于一切 $n \geqslant 1, X_n = (\xi, \mathscr{F}_n)$, (即 $X = (X_n, \mathscr{F}_n)$ 是列维鞅). 并且当 $n \to \infty$ 时 $X_n \to (\xi, \mathscr{F}_n)$ (以概率 1 收敛且在 L^1 中的收敛).*

事实上, 如果 $X = (X_n, \mathscr{F}_n)$ 是一致可积鞅, 则根据定理 2 存在可积随机变量 X_∞, 使 $X_n \to X_\infty$ (以概率 1 收敛且在 L^1 中的收敛), 并且 $X_n = (X_\infty|\mathscr{F}_n)$. 这样, 可以将 ($\mathscr{F}_\infty-$ 可测) 随机变量 ξ 取为 X_∞.

由定理 3 可得逆命题.

4. 应用列维定理的例 下面举几个列维定理应用的例.

例 1 “0–1” 律 设 $\xi_1, \xi_2, \cdots$ 是独立随机变量序列, 而 $\mathscr{F}_n^\xi = \sigma(\xi_1, \cdots, \xi_n), \mathscr{X}$ 是 “尾部” 事件的 $\sigma-$ 代数, 而 $A \in \mathscr{X}$. 由定理 3 可见

$$\mathbf{E}(I_A|\mathscr{F}_n^\xi) \to \mathbf{E}(I_A|\mathscr{F}_\infty^\xi) = I_A \quad (\mathbf{P} - \text{a.c.}).$$

由于 I_A 与 $(\xi_1, \cdots, \xi_n)$ 独立, 可见 $\mathbf{E}(I_A|\mathscr{F}_n^\xi) = \mathbf{E}I_A$, 因而 $I_A = \mathbf{E}I_A(\mathbf{P} - \text{a.c.})$, 故 $\mathbf{P}(A) = 0$ 或 $\mathbf{P}(A) = 1$.

下面的两个例子, 说明在数学分析中上面引进的关于收敛性的定理.

例 2 如果 $f = f(x)$ 是区间 $[0, 1)$ 上的函数, 并且满足利普希茨 (R. O. S. Lipschitz) 条件, 则 $f = f(x)$ 绝对连续, 并且由数学分析熟知, 存在一勒贝格可积函

数 $g = g(x)$, 使

$$f(x) - f(0) = \int_0^x g(y)\mathrm{d}y. \tag{10}$$

(在这种意义上 $g(x)$ 是 $f(x)$ 的 "导数").

现在说明由定理 1 可能得出这一结果. 设 $\Omega = [0,1), \mathscr{F} = \mathscr{B}([0,1))$, 而 $\mathbf{P}$ 是勒贝格测度. 记

$$\xi_n(x) = \sum_{k=1}^{2^n} \frac{k-1}{2^n} I\left\{\frac{k-1}{2^n} \leqslant x < \frac{k}{2^n}\right\},$$

$\mathscr{F}_n = \sigma(\xi_1, \cdots, \xi_n) = \sigma\{\xi_n\}$ 而设

$$X_n = \frac{f(\xi_n + 2^{-n}) - f(\xi_n)}{2^{-n}}.$$

因为, 对于给定的 ξ_n 的值, 随机变量 ξ_{n+1} 只有 ξ_n 和 $\xi_n + 2^{-(n+1)}$ 两个可能值, 且相应的条件概率都是 1/2, 所以

$$\begin{aligned}
&\mathbf{E}[X_{n+1}|\mathscr{F}_n] = \mathbf{E}(X_{n+1}|\xi_n) = 2^{n+1}\mathbf{E}[f(\xi_{n+1} + 2^{-(n+1)}) - f(\xi_{n+1})|\xi_n] \\
&= 2^{n+1}\left\{\frac{1}{2}[f(\xi_n + 2^{-(n+1)}) - f(\xi_n)] + \frac{1}{2}[f(\xi_n + 2^{-n}) - f(\xi_n + 2^{-(n+1)})]\right\} \\
&= 2^n[f(\xi_n + 2^{-n}) - f(\xi_n)] = X_n.
\end{aligned}$$

由此可见, $X = (X_n, \mathscr{F}_n)$ 是鞅, 并且由于 $|X_n| \leqslant L$, 其中 L 是利普希茨条件的常数: $|f(x) - f(y)| \leqslant L|x-y|$. 注意, $\mathscr{F} = \mathscr{B}([0,1)) = \sigma(\bigcup \mathscr{F}_n)$. 因此, 根据定理 3 的系, 存在 $\mathscr{F}-$ 可测函数 $g = g(x)$, 使 $X_n \to g$ ($\mathbf{P}$ − a.c.), 并且

$$X_n = \mathbf{E}(g|\mathscr{F}_n). \tag{11}$$

取集合 $B = [0, k/2^n]$, 则由 (11) 式, 有

$$f\left(\frac{k}{2^n}\right) - f(0) = \int_0^{k/2^n} x_n \mathrm{d}x = \int_0^{k/2^n} g(x)\mathrm{d}x.$$

从而, 因为 n 和 k 的任意性, 由此得所要求证明的等式 (10).

例 3　设 $\Omega = [0,1), \mathscr{F} = \mathscr{B}([0,1))$, 而 $\mathbf{P}$ 是勒贝格测度. 考虑哈尔 (A. Haar) 函数系 (由第二章 §11 例 3 的定义). 设 $\mathscr{F}_n = \sigma(H_1, \cdots, H_n)$, 并注意到 $\mathscr{F} = \mathscr{B}([0,1)) = \sigma(\bigcup \mathscr{F}_n)$. 由条件数学期望的性质以及哈尔函数的构造, 不难导出, 对于博雷尔函数 $f \in L$, 有

$$\mathbf{E}[f(x)|\mathscr{F}_n] = \sum_{k=1}^n \alpha_k H_k(x) \quad (\mathbf{P} - \text{a.c.}), \tag{12}$$

其中

$$\alpha_k = (f, H_k) = \int_0^1 f(x)H_k(x)\mathrm{d}x.$$

换句话说, 在函数 $f(x)$ 按哈尔函数系的展开时, 条件数学期望 $\mathbf{E}[f(x)|\mathscr{F}_n]$ 是傅里叶级数的部分和. 那么, 将定理 3 用于鞅 $(\mathbf{E}(f|\mathscr{F}_n),\mathscr{F}_n)$ 可得, 当 $n\to\infty$ 时

$$\sum_{k=1}^{n}(f,H_k)H_k(x)\to f(x)\ (\mathbf{P}-\text{a.c.}),$$

和

$$\int_0^1\left|\sum_{k=1}^{n}(f,H_k)H_k(x)-f(x)\right|\mathrm{d}x\to 0.$$

例 4 设 $(\xi_n)_{n\geqslant 1}$ 是随机变量序列. 根据第二章 §10 的定理 2, 如果级数 $\sum\xi_n$ 以概率 1 收敛, 则它也依概率收敛和按分布收敛. 结果表明, 假如随机变量 $\xi_1,\xi_2,\cdots$ 独立, 则逆命题也成立: 若由独立随机变量 $\xi_1,\xi_2,\cdots$ 构成的级数 $\sum\xi_n$ 按分布收敛, 则它也依概率收敛和以概率 1 收敛.

这一性质可以用如下方式证明. 设 $S_n=\xi_1+\cdots+\xi_n, n\geqslant 1$, 且 $S_n\xrightarrow{d}S$, 则对于每一个实数 $t,\mathbf{E}\mathrm{e}^{itS_n}\to\mathbf{E}\mathrm{e}^{itS}$, 显然, 存在 $\delta>0$, 使对于所有 $|t|<\delta$, 有 $|\mathbf{E}\mathrm{e}^{itS}|>0$. 取某一 t_0, 使之满足 $|t_0|<\delta$, 则存在 $n_0=n_0(t_0)$, 使得对于一切 $n\geqslant n_0$, 有 $|\mathbf{E}\mathrm{e}^{it_0S_n}|\geqslant c>0$, 其中 c 是某一常数.

对于 $n\geqslant n_0$, 建立一序列 $X=(X_n,\mathscr{F}_n)$, 其中

$$X_n=\frac{\mathrm{e}^{it_0S_n}}{\mathbf{E}\mathrm{e}^{it_0S_n}},\quad \mathscr{F}_n=\sigma(\xi_1,\cdots,\xi_n).$$

由于假设随机变量 $\xi_1,\xi_2,\cdots$ 独立, 则序列 $X=(X_n,\mathscr{F}_n)$ 是鞅, 且

$$\sup_{n\geqslant n_0}\mathbf{E}|X_n|\leqslant c^{-1}<\infty.$$

那么, 由定理 1 可见, 以概率 1 极限 $\lim\limits_n X_n$ 存在并且有限. 因此极限 $\lim\limits_n\mathrm{e}^{it_0S_n}$ 也以概率 1 存在. 因而可以断定: 存在 $\delta>0$, 使得对于集合 $T=\{t:|t|<\delta\}$ 中每一个 t, 极限 $\lim\limits_n\mathrm{e}^{itS_n}$ 以概率 1 存在.

设 $T\times\Omega=\{(t,\omega):t\in T,\omega\in\Omega\}$, $\overline{\mathscr{B}}(T)$, 是 T 上勒贝格集合的 $\sigma-$ 代数, 而 λ 是在 $(T,\overline{\mathscr{B}}(T))$ 上的勒贝格测度. 其次, 记

$$C=\{(t,\omega)\in T\times\Omega:\lim_n\mathrm{e}^{itS_n(\omega)}\ \text{存在}\}.$$

显然, $C\in\overline{\mathscr{B}}(T)\otimes\mathscr{F}$.

上面曾经证明, 对于每一个 $t\in T,\mathbf{P}(C_t)=1$, 其中 $C_t=\{\omega\in\Omega:(t,\omega)\in C\}$ 是在点 t 处集合 C 的截线. 根据傅比尼定理 (第二章 §6 的定理 8):

$$\int_{T\times\Omega}I_C(t,\omega)\mathrm{d}(\lambda\times\mathbf{P})=\int_T\left(\int_\Omega I_C(t,\omega)\mathrm{d}\mathbf{P}\right)\mathrm{d}\lambda=\int_T\mathbf{P}(C_t)\mathrm{d}\lambda=\lambda(T)=2\delta>0.$$

另一方面, 仍然根据傅比尼定理:

$$\lambda(T) = \int_{T\times\Omega} I_C(t,\omega)\mathrm{d}(\lambda\times\mathbf{P}) = \int_\Omega \left(\int_T I_C(t,\omega)\mathrm{d}\lambda\right)\mathrm{d}\mathbf{P} = \int_\Omega \lambda(C_\omega)\mathrm{d}\mathbf{P},$$

其中 $C_\omega = \{t : (t,\omega)\in C\}$.

由此可见, 存在集合 $\widetilde{\Omega}$, 且 $\mathbf{P}(\widetilde{\Omega}) = 1$, 使得对于所有 $\omega\in\widetilde{\Omega}$, 有 $\lambda(C_\omega) = \lambda(T) = 2\delta > 0$.

从而, 可以断定: 对于每一个 $\omega\in\widetilde{\Omega}$ 和对于所有 $t\in C_\omega$, 极限 $\lim\limits_n \mathrm{e}^{itS_n}$ 存在; 并且集合 C_ω 的勒贝格测度大于 0. 由此以及练习题 8, 对于每一个 $\omega\in\widetilde{\Omega}$, 极限 $\lim\limits_n S_n$ 存在且有限. 由于 $\mathbf{P}(\widetilde{\Omega}) = 1$, 可见极限 $\lim\limits_n S_n$ 以概率 1 存在并且有限.

5. 练习题

1. 设 $\{\mathscr{G}_n\}$ 是不增 $\sigma-$ 代数族, $\mathscr{G}_1\supseteq\mathscr{G}_2\supseteq\cdots$, $\mathscr{G}_\infty = \bigcap\mathscr{G}_n$, 而 η 是某一可积随机变量. 证明定理 3 之如下的类似成立: 当 $n\to\infty$ 时, $\mathbf{P}-$ 几乎处处, 以及在 L^1 收敛的意义上, 有

$$\mathbf{E}(\eta|\mathscr{G}_n)\to\mathbf{E}(\eta|\mathscr{G}_\infty).$$

2. 设 $\xi_1,\xi_2,\cdots$ 是独立同分布随机变量序列, $\mathbf{E}|\xi_1|<\infty, \mathbf{E}\xi_1 = m, S_n = \xi_1+\cdots+\xi_n$. 首先证明 (见第二章 §7 练习题 2)

$$\mathbf{E}(\xi_1|S_n,S_{n+1},\cdots) = \mathbf{E}(\xi_1|S_n) = \frac{S_n}{n}\quad (\mathbf{P}-\text{a.c.}),$$

然后利用练习题 1 的结果证明强大数定律: 当 $n\to\infty$ 时, $\mathbf{P}-$ 几乎处处以及在 L^1 的意义上, 有

$$\frac{S_n}{n}\to m.$$

3. 证明如下反映 "勒贝格控制收敛定理与 P. 列维定理" 联系的结果. 假设 $(\xi_n)_{n\geqslant 1}$ 是随机变量序列, 满足: $\xi_n\to\xi(\mathbf{P}-\text{a.c.}), |\xi_n|\leqslant\eta, \mathbf{E}\eta<\infty$, 而 $(\mathscr{F}_m)_{m\geqslant 1}$ 是非降 $\sigma-$ 代数族, $\mathscr{F}_\infty = \sigma(\bigcup\mathscr{F}_m)$. 那么, $(\mathbf{P}-\text{a.c.})$

$$\lim_{\substack{m\to\infty\\ n\to\infty}}\mathbf{E}(\xi_n|\mathscr{F}_m) = \mathbf{E}(\xi|\mathscr{F}_\infty).$$

4. 证明 (12) 式.

5. 设 $\Omega = [0,1), \mathscr{F} = \mathscr{B}([0,1))$, $\mathbf{P}$ 是勒贝格测度, 而 $f = f(x)\in L^1$. 记

$$f_n(x) = 2^n\int_{k2^{-n}}^{(k+1)2^{-n}} f(y)\mathrm{d}y,\quad k2^{-n}\leqslant x<(k+1)2^{-n}.$$

证明 $f_n(x)\to f(x)(\mathbf{P}-\text{a.c.})$.

6. 设 $\Omega = [0,1), \mathscr{F} = \mathscr{B}([0,1))$,$\mathbf{P}$ 是勒贝格测度, 而 $f = f(x) \in L^1$. 假设该函数在 $[0,2)$ 上是周期的, 并设

$$f_n(x) = \sum_{i=1}^{2^n} 2^{-n} f(x + i2^{-n}).$$

证明 $f_n(x) \to f(x)$ ($\mathbf{P}$ – a.c.).

7. 证明, 如果 $X = (X_n, \mathscr{F}_n)$ 是广义下鞅 (§1 定义 2), 满足

$$\inf_n \sup_{n \geqslant m} \mathbf{E}(X_n^+ | \mathscr{F}_m) < \infty \ (\mathbf{P} - \text{a.c.}),$$

则对于 $X = (X_n, \mathscr{F}_n)$ 定理 1 仍然成立,

8. 设 $(a_n)_{n \geqslant 1}$ 是一数列, 对于一切实数 $t : |t| < \delta (\delta > 0)$, 存在极限 $\lim\limits_n \mathrm{e}^{ita_n}$. 证明极限 $\lim\limits_n a_n$ 存在并且有限.

9. 设 $F = F(x), x \in \mathbb{R}$, 是分布函数, 而 $\alpha \in (0,1)$; 假设存在 $\theta \in \mathbb{R}$, 使 $F(\theta) = \alpha$. 构造序列 $X_1, X_2, \cdots$ (罗宾斯 – 门罗 [H. Robbins-Monroe] 方法), 使

$$X_{n+1} = X_n - n^{-1}(Y_n - \alpha),$$

其中 $Y_1, Y_2, \cdots$ 是满足

$$\mathbf{P}(Y_n = y | X_1, \cdots, X_n; Y_1, \cdots, Y_{n-1}) = \begin{cases} F(X_n), & \text{若 } y = 1, \\ 1 - F(X_n), & \text{若 } y = 0, \end{cases}$$

的随机变量.

证明 “随机逼近” 理论的如下结果: $\mathbf{E}|X_n - \theta|^2 \to 0, n \to \infty$.

10. 设 $X = (X_n, \mathscr{F}_n)_{n \geqslant 1}$ 是下鞅, 且对于每一个停时 τ, 有 $\mathbf{E}(X_\tau I(\tau < \infty)) \neq \infty$. 证明以概率 1 极限 $\lim\limits_n X_n$ 存在.

11. 设 $X = (X_n, \mathscr{F}_n)_{n \geqslant 1}$ 是鞅, 而 $\mathscr{F}_\infty = \sigma\left(\bigcup\limits_{n=1}^{\infty} \mathscr{F}_n\right)$. 证明, 如果序列 $(X_n)_{n \geqslant 1}$ 一致可积, 则极限 $X_\infty = \lim\limits_n X_n (\mathbf{P} - \text{a.c.})$ 存在, 且 “封闭” 序列 $\overline{X} = (X_n, \mathscr{F}_n)_{1 \leqslant n \leqslant \infty}$ 是鞅.

12. 设 $X = (X_n, \mathscr{F}_n)_{n \geqslant 1}$ 是下鞅, 而 $\mathscr{F}_\infty = \sigma(\bigcup\limits_{n=1}^{\infty} \mathscr{F}_n)$. 证明, 如果序列 $(X_n^+)_{n \geqslant 1}$ 一致可积, 则极限 $X_\infty = \lim\limits_n X_n (\mathbf{P} - \text{a.c.})$ 存在, 且 “封闭” 序列 $\overline{X} = (X_n, \mathscr{F}_n)_{1 \leqslant n < \infty}$ 是下鞅.

§5. 下鞅和鞅的收敛集

1. 随机序列类 C^+ 设 $X = (X_n, \mathscr{F}_n)$ 是随机序列以 $\{X_n \to\}$ 或 $\{-\infty < \lim X_n < \infty\}$ 表示 “使 $\lim X_n$ *存在和有限*的基本结局的集合”. 对于任意二事件 A 和 B, 如果 $\mathbf{P}\{I_A \leqslant I_B\} = 1$, 则亦称 $A \subseteq B (\mathbf{P} - \text{a.c.})$.

如果 X 是下鞅且 $\sup \mathbf{E}|X_n| < \infty$ (或者等价地 $\sup \mathbf{E}X_n^+ < \infty$), 则根据 §4 定理 1, 有

$$\{X_n \to\} = \Omega \quad (\mathbf{P}-\text{a.c.}), \quad \text{即 } \{X_n \nrightarrow\} = 0.$$

现在讨论, 在条件 $\sup \mathbf{E}|X_n| < \infty$ 不成立的情况下, 关于下鞅的收敛集合 $\{X_n \to\}$ 的结构问题.

设 $a > 0$,

$$\tau_a = \begin{cases} \inf\{n \geqslant 1 : X_n > a\}, & \text{若 } \{\cdot\} \neq \varnothing, \\ \infty, & \text{若 } \{\cdot\} = \varnothing. \end{cases}$$

定义　如果对于任意 $a > 0$, 有

$$\mathbf{E}(\Delta X_{\tau_a})^+ I(\tau_a < \infty) < \infty, \tag{1}$$

则称随机序列 $X = (X_n, \mathscr{F}_n)$ 属于类 $\mathbf{C}^+ (X \in \mathbf{C}^+)$, 其中 $\Delta X_n = X_n - X_{n-1}, X_0 = 0$.

显然, 如果

$$\mathbf{E} \sup_n |\Delta X_n| < \infty \tag{2}$$

则 $X \in \mathbf{C}^+$; 特别, 如果 $(\mathbf{P}-\text{a.c.})$ 对于一切 $n \geqslant 1$,

$$|\Delta X_n| \leqslant C < \infty. \tag{3}$$

定理 1　如果下鞅 $X \in \mathbf{C}^+$, 则 $(\mathbf{P}-\text{a.c.})$

$$\{\sup X_n < \infty\} = \{X_n \to\}. \tag{4}$$

证明　包含关系 $\{X_n \to\} \subseteq \{\sup X_n < \infty\}$ 显然. 为证明相反的包含关系, 考虑 "停止" 下鞅 $X^{\tau_a} = (X_{\tau_a \wedge n}, \mathscr{F}_n)$. 那么, 由于 (1) 式, 有

$$\sup_n \mathbf{E}X_{\tau_a \wedge n}^+ \leqslant a + \mathbf{E}[X_{\tau_a}^+ I\{\tau_a < \infty\}] \leqslant 2a + \mathbf{E}[(\Delta X_{\tau_a})^+ I\{\tau_a < \infty\}] < \infty. \tag{5}$$

从而, 由 §4 定理 1 $(\mathbf{P}-\text{a.c.})$, 有

$$\{\tau_a = \infty\} \subseteq \{X_n \to\}.$$

由于

$$\bigcup_{a>0} \{\tau_a = \infty\} = \{\sup X_n < \infty\},$$

可见 $\{\sup X_n < \infty\} \subseteq \{X_n \to\} (\mathbf{P}-\text{a.c.})$. □

系 设 X 是鞅, 且 $\mathbf{E}\sup|\Delta X_n| < \infty$. 那么,

$$\{X_n \to\} \bigcup \{\underline{\lim} X_n = -\infty, \overline{\lim} X_n = +\infty\} = \Omega \ (\mathrm{P-a.c.}). \tag{6}$$

实际上, 将定理 1 用于 X 和 $-X$, 可得

$$\left\{\overline{\lim} X_n < \infty\right\} = \{\sup X_n < \infty\} = \{X_n \to\} \ (\mathrm{P-a.c.}),$$
$$\{\underline{\lim} X_n > -\infty\} = \{\inf X_n > -\infty\} = \{X_n \to\} \ (\mathrm{P-a.c.}).$$

因此,

$$\left\{\overline{\lim} X_n < \infty\right\} \bigcup \ \{\underline{\lim} X_n > -\infty\} = \{X_n \to\} \ (\mathrm{P-a.c.}),$$

于是, (6) 式得证,

命题 (6) 表示, 对于满足条件 $\mathbf{E}\sup|\Delta X_n| < \infty$ 的鞅 X 的几乎一切轨道, 或者存在有限极限, 或者这些轨道按如下意义上, 表现 "不好": $\overline{\lim} X_n = +\infty, \underline{\lim} X_n = -\infty$.

2. C^+ 类非负鞅的性质 如果 $\xi_1, \xi_2, \cdots$ 是独立随机变量序列, 且 $\mathbf{E}\xi_i = 0, |\xi_i| \leqslant c < \infty$, 则根据第四章 §2 定理 1, 级数 $\sum \xi_i$ 收敛 ($\mathbf{P}$ − a.c.), 当且仅当 $\sum \xi_i^2 < \infty$. 序列

$$X = (X_n, \mathscr{F}_n), \quad X_n = \xi_1 + \cdots + \xi_n, \quad \mathscr{F}_n = \sigma\{\xi_1, \cdots, \xi_n\},$$

是平方可积鞅, 且 $\langle X\rangle_n = \sum\limits_{i=1}^{n} \xi_i^2$, 而可以将命题表述为如下形式:

$$\{\langle X\rangle_\infty < \infty\} = \{X_n \to\} = \Omega, \quad (\mathbf{P} - \mathrm{a.c.}),$$

其中 $\langle X\rangle_\infty = \lim\limits_n \langle X\rangle_n$.

下面的命题将这一结果推广到鞅和下鞅的更一般的情形.

定理 2 设 $X = (X_n, \mathscr{F}_n)$ 是下鞅, 而

$$X_n = m_n + A_n$$

是其杜布分解.

a) 如果 X 是非负下鞅, 则

$$\{A_\infty < \infty\} \subseteq \{X_n \to\} \subseteq \{\sup X_n < \infty\} \ (\mathbf{P} - \mathrm{a.c.}). \tag{7}$$

b) 如果 $X \in \mathbf{C}^+$, 则

$$\{X_n \to\} = \{\sup X_n < \infty\} \subseteq \{A_\infty < \infty\} \ (\mathbf{P} - \mathrm{a.c.}). \tag{8}$$

c) 如果 X 是非负下鞅, 且 $X \in \mathbf{C}^+$, 则

$$\{X_n \to\} = \{\sup X_n < \infty\} = \{A_\infty < \infty\} \ (\mathbf{P} - \mathrm{a.c.}). \tag{9}$$

证明 a) (7) 式的第二个包含关系显然. 为证明 (7) 式的第一个包含关系, 引进时间

$$\sigma_a = \begin{cases} \inf\{n \geqslant 1 : A_{n+1} > a\}, & \text{若 } a > 0, \\ +\infty, & \text{若 } \{\cdot\} = \varnothing. \end{cases}$$

那么, $A_{\sigma_a} \leqslant a$, 且由于 §2 定理 1 的系 1, 可见

$$\mathbf{E}X_{n\wedge\sigma_a} = \mathbf{E}A_{n\wedge\sigma_a} \leqslant a.$$

设 $Y_n^a = X_{n\wedge\sigma_a}$, 则 $Y^a = (Y_n^a, \mathscr{F}_n)$ 是下鞅, 其中 $\sup \mathbf{E}Y_n^a \leqslant a < \infty$, 而由于其非负性, 由 §4 定理 1, 可见.

$$\{A_\infty < \infty\} = \{\sigma_a = \infty\} \subseteq \{X_n \to\} \ (\mathbf{P} - \text{a.c.}).$$

因此

$$\{A_\infty < \infty\} = \bigcup_{a>0} \{A_\infty \leqslant a\} \subseteq \{X_n \to\} \ (\mathbf{P} - \text{a.c.}).$$

b) 由定理 1 可得 (8) 式的第一个等式. 为证明 (8) 式中的包含关系, 注意到, 根据关系式 (5), 有

$$\mathbf{E}A_{\tau_a\wedge n} = \mathbf{E}X_{\tau_a\wedge n} \leqslant \mathbf{E}X^+_{\tau_a\wedge n} \leqslant 2a + \mathbf{E}[(\Delta X_{\tau_a})^+ I\{\tau_a < \infty\}],$$

从而

$$\mathbf{E}A_{\tau_a} = \mathbf{E}\lim_n A_{\tau_a\wedge n} < \infty.$$

于是, 由

$$\bigcup_{a>0} \{\tau_a = \infty\} = \{\sup X_n < \infty\},$$

得 $\{\tau_a = \infty\} \subseteq \{A_\infty < \infty\}$ 和所要证明的 (8) 式.

c) 该命题是命题 a) 和 b) 的直接推论. □

注 可以将 "X 的非负性" 条件换成 $\sup\limits_n \mathbf{E}X_n^- < \infty$.

系 1 设 $X_n = \xi_1 + \cdots + \xi_n$, 其中 $\xi_i \geqslant 0, \mathbf{E}\xi_i < \infty$, 且 ξ_i 为 $\mathscr{F}_i$– 可测的, 而 $\mathscr{F}_0 = \{\varnothing, \Omega\}$. 那么, ($\mathbf{P}$–a.c.)

$$\left\{\sum_{n=1}^{\infty} \mathbf{E}(\xi_n|\mathscr{F}_{n-1}) < \infty\right\} \subseteq \{X_n \to\}, \tag{10}$$

并且. 如果 $\mathbf{E}\sup\limits_n \xi_n < \infty$, 则($\mathbf{P}$–a.c.)

$$\left\{\sum_{n=1}^{\infty} \mathbf{E}(\xi_n|\mathscr{F}_{n-1}) < \infty\right\} = \{X_n \to\}. \tag{11}$$

系 2 (博雷尔 – 坎泰利 – 列维引理) 如果事件 $B_n \in \mathscr{F}_n$, 则在 (11) 式中设 $\xi_n = I_{B_n}$, 得 ($\mathbf{P}$–a.c.)

$$\left\{\sum_{n=1}^{\infty} \mathbf{P}(B_n|\mathscr{F}_{n-1}) < \infty\right\} = \left\{\sum_{n=1}^{\infty} I_{B_n} < \infty\right\}. \tag{12}$$

3. 平方可积鞅的性质

定理 3 设 $M = (M_n, \mathscr{F}_n)_{n\geqslant 1}$ 是平方可积鞅, 则 ($\mathbf{P}$–a.c.)

$$\{\langle M\rangle_\infty < \infty\} \subseteq \{M_n \to\}. \tag{13}$$

假如还满足 $\mathbf{E}\sup|\Delta M_n|^2 < \infty$, 那么 ($\mathbf{P}$–a.c.)

$$\{\langle M\rangle_\infty < \infty\} = \{M_n \to\}, \tag{14}$$

其中

$$\langle M\rangle_\infty = \sum_{n=1}^{\infty} \mathbf{E}[(\Delta M_n)^2|\mathscr{F}_{n-1}], \tag{15}$$

而 $M_0 = 0$, $\mathscr{F}_0 = \{\varnothing, \Omega\}$.

证明 考虑两个下鞅: $M^2 = (M_n^2, \mathscr{F}_n)$ 和 $(M+1)^2 = ((M+1)_n^2, \mathscr{F}_n)$. 那么, 在它们的杜布分解

$$M^2 = m_n' + A_n' \quad 和 \quad (M+1)^2 = m_n'' + A_n''$$

中随机变量 A_n' 和 A_n'' 相等, 因为

$$A_n' = \sum_{k=1}^{n} \mathbf{E}(\Delta M_n^2|\mathscr{F}_{k-1})$$

而

$$A_n'' = \sum_{k=1}^{n} \mathbf{E}[\Delta(M_k+1)^2|\mathscr{F}_{k-1}] = \sum_{k=1}^{n} \mathbf{E}(\Delta M_n^2|\mathscr{F}_{k-1}) = A_n'.$$

从而, 由 (7) 式, 以概率 1 有

$$\{\langle M\rangle_\infty < \infty\} = \{A_\infty' < \infty\} \subseteq \{M_n^2 \to\} \bigcap \{(M_n+1)^2 \to\} = \{M_n \to\}.$$

由于 (9) 式, 为证明 (14) 式只需验证, 条件 $\mathbf{E}\sup|\Delta M_n|^2 < \infty$ 可以保障下鞅 M^2 属于 $\mathbf{C}^+$ 类: $M^2 \in \mathbf{C}^+$.

设 $\tau_a = \inf\{n \geqslant 1 : M_n^2 > a\}, a > 0$, 则在集合 $\{\tau_a < \infty\}$ 上, 有

$$\begin{aligned}|M_{\tau_a}^2| = |M_{\tau_a}^2 - M_{\tau_a-1}^2| &\leqslant |M_{\tau_a} - M_{\tau_a-1}|^2 + 2|M_{\tau_a-1}| \cdot |M_{\tau_a} - M_{\tau_a-1}| \\ &\leqslant (\Delta M_{\tau_a})^2 + 2a^{1/2}|\Delta M_{\tau_a}|,\end{aligned}$$

由此, 得

$$\begin{aligned}\mathbf{E}|M_{\tau_a}^2|I\{\tau_a<\infty\} &\leqslant \mathbf{E}(\Delta M_{\tau_a})^2 I\{\tau_a<\infty\}+2a^{1/2}\sqrt{\mathbf{E}(\Delta M_{\tau_a})^2 I\{\tau_a<\infty\}}\\ &\leqslant \mathbf{E}\sup|\Delta M_n|^2+2a^{1/2}\sqrt{\mathbf{E}\sup|\Delta M_n|^2}<\infty.\end{aligned}$$ □

作为应用这一定理的例子, 我们引进下面的结果, 可以视为平方可积鞅的强大数定律的独特形式 (对照第四章 §3 定理 2, 以及第四章第 3 小节系 2).

定理 4 *设 $M=(M_n,\mathscr{F}_n)$ 是平方可积鞅, 而 $A=(A_n,\mathscr{F}_{n-1})$ 是可预测的递增序列, 且 $A_1\geqslant 1, A_\infty=\infty(\mathbf{P}-\text{a.c.})$.*

如果 $(\mathbf{P}-\text{a.c.})$

$$\sum_{i=1}^{\infty}\frac{\mathbf{E}[(\Delta M_i)^2|\mathscr{F}_{i-1}]}{A_i^2}<\infty, \tag{16}$$

则以概率 1, 有

$$\frac{M_n}{A_n}\to 0,\quad n\to\infty. \tag{17}$$

特别, 如果 $\langle M\rangle=(\langle M\rangle_n,\mathscr{F}_{n-1})$ 是平方可积鞅 $M=(M_n,\mathscr{F}_n)$ 的二次特征, 且 $\langle M\rangle_\infty=\infty(\mathbf{P}-\text{a.c.})$, 则以概率 1, 有

$$\frac{M_n}{\langle M\rangle_n}\to 0,\quad n\to\infty. \tag{18}$$

证明 考虑平方可积鞅 $m=(m_n,\mathscr{F}_n)$, 且

$$m_n=\sum_{i=1}^{n}\frac{\Delta M_i}{A_i}.$$

那么,

$$\langle m\rangle_n=\sum_{k=1}^{n}\frac{\mathbf{E}[(\Delta M_i)^2|\mathscr{F}_{i-1}]}{A_i^2}. \tag{19}$$

由于

$$\frac{M_n}{A_n}=\frac{\sum\limits_{k=1}^{n}A_k\Delta m_k}{A_n},$$

可见, 根据克罗内克引理 (第四章 §3), 如果以概率 1 存在有限极限 $\lim m_n$, 则

$$\frac{M_n}{A_n}\to 0,\quad n\to\infty\ (\mathbf{P}-\text{a.c.}).$$

而由于 (13) 式, 有

$$\{\langle m\rangle_\infty<\infty\}\leqslant\{m_n\to\}, \tag{20}$$

因此由 (19) 式可见, (16) 式是 (17) 式成立的充分条件.

最后, 假如 $A_n=\langle M\rangle_n$, 则条件 (16) 自然成立 (练习题 6). □

例 考虑独立随机变量序列 $\xi_1, \xi_2, \cdots$, 且 $\mathbf{E}\xi_i = 0, \mathbf{D}\xi_i = D_i > 0$, 而序列 $X = (X_n)_{n\geqslant 0}$ 决定于如下递推方程:

$$X_{n+1} = \theta X_n + \xi_{n+1}, \tag{21}$$

其中 X_0 不依赖于 $\xi_1, \xi_2 \cdots$, 而 $\theta(-\infty < \theta < \infty)$ 是未知参数.

我们把 X_n 视为在时刻 n 的观测结果, 并且要求估计未知参数 θ. 以量

$$\widehat{\theta}_n = \frac{\sum_{k=0}^{n-1} \frac{X_k X_{k+1}}{D_{k+1}}}{\sum_{k=0}^{n-1} \frac{X_k^2}{D_{k+1}}}, \tag{22}$$

为根据观测结果 $X_0, X_1, \cdots, X_n$, 对 θ 的估计, 且当分母为 0 时, 设 $\widehat{\theta}_n = 0$. ($\widehat{\theta}_n$ 是由最小二乘法得来的估计.)

由 (21) 和 (22) 式, 显然有

$$\widehat{\theta}_n = \theta + \frac{M_n}{A_n},$$

其中

$$M_n = \sum_{k=0}^{n-1} \frac{X_k \xi_{k+1}}{D_{k+1}}, \quad A_n = \langle M \rangle_n = \sum_{k=0}^{n-1} \frac{X_k^2}{D_{k+1}}.$$

因此, 假如未知参数的真值为 θ, 则

$$\mathbf{P}\{\widehat{\theta}_n \to \theta\} = 1 \tag{23}$$

当且仅当 ($\mathbf{P}$ – a.c.)

$$\frac{M_n}{A_n} \to 0, \quad n \to \infty. \tag{24}$$

现在证明, 条件

$$\sup_n \frac{D_{n+1}}{D_n} < \infty, \quad \sum_{n=1}^{\infty} \left(\frac{\xi_n^2}{D_n} \wedge 1 \right) = \infty, \tag{25}$$

对于 (24) 式是充分的, 因此对于 (23) 式也是充分的. 有

$$\begin{aligned} \sum_{n=1}^{\infty} \left(\frac{\xi_n^2}{D_n} \wedge 1 \right) &\leqslant \sum_{n=1}^{\infty} \frac{\xi_n^2}{D_n} = \sum_{n=1}^{\infty} \frac{(X_n - \theta X_{n-1})^2}{D_n} \\ &\leqslant 2 \left[\sum_{n=1}^{\infty} \frac{X_n^2}{D_n} + \theta^2 \sum_{n=1}^{\infty} \frac{X_{n-1}^2}{D_n} \right] \leqslant 2 \left[\sup \frac{D_{n+1}}{D_n} + \theta^2 \right] \langle M \rangle_\infty. \end{aligned}$$

于是,

$$\left\{ \sum_{n=1}^{\infty} \left(\frac{\xi_n^2}{D_n} \wedge 1 \right) = \infty \right\} \subseteq \{ \langle M \rangle_\infty = \infty \}.$$

根据“三级数”定理 (第四章 §2 定理 3), 若级数

$$\sum_{n=1}^{\infty}\mathbf{E}\left(\frac{\xi_n^2}{D_n}\wedge 1\right)$$

发散, 则级数

$$\sum_{n=1}^{\infty}\left(\frac{\xi_n^2}{D_n}\wedge 1\right),$$

以概率 1 也发散. 从而, $\mathbf{P}\{\langle M\rangle_\infty=\infty\}=1$, 故由定理 4 的论断 (7) 直接得 (24) 式.

具有性质 (23) 的估计量 $\widehat{\theta}_n, n\geqslant 1$, 称做强相合的 (对照第一章 §7 中“相合性”的概念).

在下一节 §6 的第 5 小节, 对于高斯序列 $\xi_1,\xi_2\cdots$ 的情形, 将继续这个例子.

定理 5 设 $X=(X_n,\mathscr{F}_n)$ 是下鞅, 而

$$X_n=m_n+A_n$$

是其杜布分解. 如果 $|\Delta X_n|\leqslant C$, 则 ($\mathbf{P}$ – a.c.)

$$\{\langle m\rangle_\infty+A_\infty<\infty\}=\{X_n\rightarrow\},\tag{26}$$

或同样地,

$$\left\{\sum_{n=1}^{\infty}\mathbf{E}[\Delta X_n+(\Delta X_n)^2|\mathscr{F}_{n-1}]\right\}=\{X_n\rightarrow\}.\tag{27}$$

证明 由于

$$A_n=\sum_{n=1}^{\infty}\mathbf{E}[(\Delta X_n)^2|\mathscr{F}_{n-1}],\tag{28}$$

$$m_n=\sum_{k=1}^{n}\mathbf{E}[\Delta X_k-(\Delta X_n|\mathscr{F}_{k-1})],\tag{29}$$

则由于假设 $|\Delta X_n|\leqslant C$, 鞅 $m=(m_n,\mathscr{F}_n)$ 平方可积, 且 $|\Delta m_n|\leqslant 2C$. 那么, 由 (13) 式, 有

$$\{\langle m\rangle_\infty+A_\infty<\infty\}\subseteq\{X_n\rightarrow\},\tag{30}$$

而且, 根据 (8) 式, 有

$$\{X_n\rightarrow\}\subseteq\{A_\infty<\infty\}.$$

因此, 由 (14) 式和 (30) 式, 可见

$$\begin{aligned}\{X_n\rightarrow\}&=\{X_n\rightarrow\}\bigcap\{A_\infty<\infty\}=\{X_n\rightarrow\}\bigcap\{A_\infty<\infty\}\bigcap\{m_n\rightarrow\}\\&=\{X_n\rightarrow\}\bigcap\{A_\infty<\infty\}\bigcap\{\langle m\rangle_\infty<\infty\}\\&=\{X_n\rightarrow\}\bigcap\{A_\infty+\langle m\rangle_\infty<\infty\}=\{A_\infty+\langle m\rangle_\infty<\infty\}.\end{aligned}$$

最后, 由于 (29) 式, 有

$$\langle m\rangle_n=\sum_{k=1}^{n}\left\{\mathbf{E}[(\Delta X_k)^2|\mathscr{F}_{k-1}]-[\mathbf{E}(\Delta X_k|\mathscr{F}_{k-1})]^2\right\};$$

可见 (26) 式和 (27) 式等价. 而由非负项级数

$$\sum_{k=1}^{\infty}\mathbf{E}(\Delta X_k|\mathscr{F}_{k-1})$$

的收敛性, 可见级数

$$\sum_{k=1}^{\infty}[\mathbf{E}(\Delta X_k|\mathscr{F}_{k-1})]^2$$

的收敛性. □

4. 级数 $\sum\xi_n$ 的收敛集合 柯尔莫戈洛夫 "三级数" 定理 (第四章 §2 定理 3) 给出了, 独立随机变量的级数 $\sum\xi_n$ 以概率 1 收敛的充分且必要条件. 下面的定理 6, 在不假设随机变量 $\xi_1,\xi_2,\cdots$ 独立的情形下, 给出了级数 $\sum\xi_n$ 之收敛集合的描述, 而其证明基于定理 2 和定理 3.

定理 6 假设 $\xi=(\xi_n,\mathscr{F}_n)_{n\geqslant1}$ 是随机变量序列, $\mathscr{F}_0=(\varnothing,\Omega)$ 而是正常数. 那么, 级数 $\sum\xi_n$ 在集合 A 上收敛, 其中 A 是使三个级数

$$\sum\mathbf{P}(|\xi_n|\geqslant c|\mathscr{F}_{n-1}),\quad \sum\mathbf{E}(\xi_n^c|\mathscr{F}_{n-1}),\quad \sum\mathbf{D}(\xi_n^c|\mathscr{F}_{n-1}),$$

同时收敛的集合, 其中 $\xi_n^c=\xi_nI(|\xi_n|\leqslant c)$.

证明 设级数 $X_n=\sum\limits_{k=1}^{n}\xi_k$. 由于级数 $\sum\mathbf{P}(|\xi_n|\geqslant c|\mathscr{F}_{n-1})$ (在集合 A 上) 收敛, 则根据定理 2 的系 2, 以及级数 $\sum\mathbf{E}(\xi_n^c|\mathscr{F}_{n-1})$ 收敛, 有

$$\begin{aligned}A\bigcap\{X_n\to\}&=A\bigcap\left\{\sum_{k=1}^{n}\xi_kI(|\xi_k|\leqslant c)\to\right\}\\&=A\bigcap\left\{\sum_{k=1}^{n}[\xi_kI(|\xi_k|\leqslant c)-\mathbf{E}(\xi_kI(|\xi_k|\leqslant c)|\mathscr{F}_{n-1})]\to\right\}.\end{aligned}\tag{31}$$

设 $\eta_k=\xi_k^c-\mathbf{E}(\xi_k^c|\mathscr{F}_{n-1})$ 和 $Y_n=\sum\limits_{k=1}^{n}\eta_k$. 那么, $Y=(Y_n,\mathscr{F}_n)$ 是平方可积鞅, 其中 $|\eta_k|\leqslant 2c$. 根据定理 3,

$$A\subseteq\left\{\sum\mathbf{D}(\xi_n^c|\mathscr{F}_{n-1})<\infty\right\}=\{\langle Y\rangle_\infty<\infty\}=\{Y_n\to\}.$$

因此, 由 (31) 可见

$$A\bigcap\{X_n\to\}=A.$$

于是, $A\subseteq\{X_n\to\}$. □

5. 练习题

1. 证明, 如果下鞅 $X=(X_n,\mathscr{F}_n)$ 满足条件 $\mathbf{E}\sup\limits_n|X_n|<\infty$, 则它属于 $\mathbf{C}^+$ 类.
2. 证明, 对于广义下鞅, 定理 1 和定理 2 仍然成立.
3. 证明, 对于广义下鞅, 以概率 1 有包含关系:

$$\left\{\inf_m \sup_{n\geqslant m}\mathbf{E}(X_n^+|\mathscr{F}_m)<\infty\right\}\subseteq\{X_n\to\}.$$

4. 证明, 对于广义鞅, 定理 1 的系仍然成立.
5. 证明, $\mathbf{C}^+$ 类的任何广义下鞅是局部下鞅.
6. 设 $a_n>0,n\geqslant 1;b_n=\sum\limits_{k=1}^{n}a_k$, 证明

$$\sum_{k=1}^{\infty}\frac{a_n}{b_n^2}<\infty.$$

7. 设 $\xi_0,\xi_1,\xi_2,\cdots$ 是一致有界随机变量序列: $|\xi_n|\leqslant c,n\geqslant 1$, 证明级数

$$\sum_{n\geqslant 0}\xi_n \quad 和 \quad \sum_{n\geqslant 1}\mathbf{E}(\xi_n|\xi_1,\cdots,\xi_{n-1}),$$

以概率 1 或者同时收敛, 或者同时发散.

§6. 概率测度在带滤子可测空间上的绝对连续性和奇异性

1. 测度的局部绝对连续性和奇异性 设 $(\Omega,\mathscr{F}_n)$ 是某一可测空间, 而 $(\mathscr{F}_n)_{n\geqslant 1}$ 是该空间上的 $\sigma-$ 代数族, 且 $\mathscr{F}_1\subseteq\subsetneq\mathscr{F}_2\subseteq\cdots\subseteq\mathscr{F}$, 其中

$$\mathscr{F}=\sigma\left(\bigcup_{n=1}^{\infty}\mathscr{F}_n\right). \tag{1}$$

假设在 $(\Omega,\mathscr{F}_n)$ 上有两个概率测度 $\mathbf{P}$ 和 $\widetilde{\mathbf{P}}$. 记

$$\mathbf{P}_n=\mathbf{P}|\mathscr{F}_n \quad \widetilde{\mathbf{P}}_n=\widetilde{\mathbf{P}}|\mathscr{F}_n$$

为两个概率测度在上 $\mathscr{F}_n$ 的收缩, 即 $\mathbf{P}_n$ 和 $\widetilde{\mathbf{P}}_n$ 是 $(\Omega,\mathscr{F}_n)$ 上的概率测度, 并且对于 $B\in\mathscr{F}_n$, 有

$$\mathbf{P}_n(B)=\mathbf{P}(B),\quad \widetilde{\mathbf{P}}_n(B)=\widetilde{\mathbf{P}}(B).$$

熟知, 测度 $\widetilde{\mathbf{P}}$ 称做关于测度 $\mathbf{P}$ 绝对连续的 (记作 $\widetilde{\mathbf{P}}\ll\mathbf{P}$), 如果 $\mathbf{P}_n(A)=0,A\in\mathscr{F}_n$, 则必有 $\widetilde{\mathbf{P}}_n(A)=0$.

如果 $\widetilde{\mathbf{P}}\ll\mathbf{P}$ 且 $\mathbf{P}\ll\widetilde{\mathbf{P}}$, 则称测度 $\mathbf{P}$ 和 $\widetilde{\mathbf{P}}$ 为等价的 (记作 $\widetilde{\mathbf{P}}\sim\mathbf{P}$).

测度 $\widetilde{\mathbf{P}}$ 和 $\mathbf{P}$ 称做奇异的或正交的 (记作 $\widetilde{\mathbf{P}}\perp\mathbf{P}$), 如果存在集合 $A\in\mathscr{F}$, 使 $\widetilde{\mathbf{P}}(A)=1$和 $\mathbf{P}(\overline{A})=1$.

定义 1 称测度 $\widetilde{\mathbf{P}}$ 关于测度 $\mathbf{P}$ 为局部绝对连续的 (记作 $\widetilde{\mathbf{P}} \overset{\text{loc}}{\ll} \mathbf{P}$), 如果对于任意 $n \geqslant 1$, 有

$$\widetilde{\mathbf{P}}_n \ll \mathbf{P}_n. \tag{2}$$

这一节考虑的基本问题在于说明, 由局部绝对连续性 $\widetilde{\mathbf{P}} \overset{\text{loc}}{\ll} \mathbf{P}$, 说明 $\widetilde{\mathbf{P}} \ll \mathbf{P}, \widetilde{\mathbf{P}} \sim \mathbf{P}, \widetilde{\mathbf{P}} \perp \mathbf{P}$ 等性质成立的条件. 我们以后将清楚地看到, 鞅论是可以全面地回答这些问题的数学工具.

回忆, 在第三章 §9 中. 对于任意概率测度, 讨论过概率测度的绝对连续性和奇异性问题. 曾经说明, 利用海林格 (E. Helinger) 积分可以表述相应的准则 (定理 2 和定理 3). 下面引进的关于概率测度的绝对连续性和奇异性的结果, 亦可由这些准则得到 (在专著 [84], [87] 中陈述了有关方法). 为更完整地演示运用 §5 中关于下鞅的收敛集合, 我们在此倾向于略有不同的叙述方法. (注意, 这一节全部叙述, 假设局部绝对连续性成立的. 这样做仅仅是为了简便. 对于一般情形, 请读者参阅 [84], [87]).

这样, 假设 $\widetilde{\mathbf{P}} \overset{\text{loc}}{\ll} \mathbf{P}$. 设

$$z_n = \frac{\mathrm{d}\widetilde{\mathbf{P}}_n}{\mathrm{d}\mathbf{P}_n}$$

是 $\widetilde{\mathbf{P}}_n$ 关于 $\mathbf{P}_n$ 的拉东 – 尼科迪姆导数. 显然, z_n 为 $\mathscr{F}_n$– 可测, 并且如果 $A \in \mathscr{F}_n$, 则

$$\int_A z_{n+1}\mathrm{d}\mathbf{P} = \int_A \frac{\mathrm{d}\widetilde{\mathbf{P}}_{n+1}}{\mathrm{d}\mathbf{P}_{n+1}}\mathrm{d}\mathbf{P} = \widetilde{\mathbf{P}}_{n+1}(A) = \widetilde{\mathbf{P}}_n(A) = \int_A \frac{\mathrm{d}\widetilde{\mathbf{P}}_n}{\mathrm{d}\mathbf{P}_n}\mathrm{d}\mathbf{P} = \int_A z_n\mathrm{d}\mathbf{P}.$$

由此可见, 随机序列 $z = (z_n, \mathscr{F}_n)_{n\geqslant 1}$ 关于测度 $\mathbf{P}$ 是鞅.

在 "绝对连续性和奇异性" 的全部问题中, 关键环节是下面的定理.

定理 1 *设* $\widetilde{\mathbf{P}} \overset{\text{loc}}{\ll} \mathbf{P}$.

a) $(\mathbf{P} + \widetilde{\mathbf{P}})/2$ *以概率 1 存在极限* $\lim\limits_n z_n$, *记为* z_∞, *使*

$$\mathbf{P}\{z_\infty = \infty\} = 0.$$

b) *有如下勒贝格分解*:

$$\widetilde{\mathbf{P}}(A) = \int_A z_\infty\mathrm{d}\mathbf{P} + \widetilde{\mathbf{P}}(A \bigcap \{z_\infty = \infty\}), \quad A \in \mathscr{F}, \tag{3}$$

而且测度 $\widetilde{\mathbf{P}}(A \bigcap \{z_\infty = \infty\})$ *和* $\mathbf{P}(A), A \in \mathscr{F}$, *是奇异的*.

证明 我们首先指出, 根据经典的勒贝格分解 (第三章 §9 的 (29) 式), 任意概率测度 $\widetilde{\mathbf{P}}$ 关于概率测度 $\mathbf{P}$ 有如下表现:

$$\widetilde{\mathbf{P}}(A) = \int_A \frac{\widetilde{\mathfrak{z}}}{\mathfrak{z}}\mathrm{d}\mathbf{P} + \widetilde{\mathbf{P}}(A \bigcap \{\mathfrak{z} = 0\}), \quad A \in \mathscr{F}, \tag{4}$$

其中

$$\mathfrak{z} = \frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}}, \quad \widetilde{\mathfrak{z}} = \frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{Q}},$$

而其中测度 $\mathbf{Q}$, 例如可以取作 $\mathbf{Q} = (\mathbf{P} + \widetilde{\mathbf{P}})/2$. 这样, (3) 式可以视为一般分解 (4) 的具体化, 与所考虑情形的如下特征相联系: $\widetilde{\mathbf{P}} \overset{\mathrm{loc}}{\ll} \mathbf{P}$ 即 $\widetilde{\mathbf{P}}_n \ll \mathbf{P}_n, n \geqslant 1$. 设

$$\mathfrak{z}_n = \frac{\mathrm{d}\mathbf{P}_n}{\mathrm{d}\mathbf{Q}_n}, \quad \widetilde{\mathfrak{z}}_n = \frac{\mathrm{d}\widetilde{\mathbf{P}}_n}{\mathrm{d}\mathbf{Q}_n}, \quad \mathbf{Q}_n = \frac{1}{2}(\mathbf{P}_n + \widetilde{\mathbf{P}}_n).$$

序列 $(\mathfrak{z}_n, \mathscr{F}_n)$ 和 $(\widetilde{\mathfrak{z}}_n, \mathscr{F}_n)$ 关于测度 $\mathbf{Q}$ 是鞅, 并且满足 $0 \leqslant \mathfrak{z}_n \leqslant 2, 0 \leqslant \widetilde{\mathfrak{z}}_n \leqslant 2$. 因此 (根据 §4 定理 2), 既关于测度 $\mathbf{Q}$ 几乎处处, 又在 $L^1(\Omega, \mathscr{F}, \mathbf{Q})$ 的收敛意义上存在极限:

$$\mathfrak{z}_\infty = \lim_n \mathfrak{z}_n, \quad \widetilde{\mathfrak{z}}_\infty = \lim_n \widetilde{\mathfrak{z}}_n. \tag{5}$$

特别, 由在 $L^1(\Omega, \mathscr{F}, \mathbf{Q})$ 的意义上收敛, 可见对于任意 $A \in \mathscr{F}_m$, 有

$$\int_A \widetilde{\mathfrak{z}}_\infty \mathrm{d}\mathbf{Q} = \lim_{n\uparrow\infty} \int_A \widetilde{\mathfrak{z}}_n \mathrm{d}\mathbf{Q} = \int_A \widetilde{\mathfrak{z}}_m \mathrm{d}\mathbf{Q} = \widetilde{\mathbf{P}}_m(A) = \widetilde{\mathbf{P}}(A).$$

那么, 由卡拉泰奥多里 (C. Carathéodory) 定理 (第二章 §3) 可见, 对于任意 $A \in \mathscr{F} = \sigma\left(\bigcup_n \mathscr{F}_n\right)$, 有

$$\int_A \widetilde{\mathfrak{z}}_\infty \mathrm{d}\mathbf{Q} = \widetilde{\mathbf{P}}(A),$$

即 $\mathrm{d}\widetilde{\mathbf{P}}/\mathrm{d}\mathbf{Q} = \widetilde{\mathfrak{z}}_\infty$; 类似地, 有

$$\int_A \mathfrak{z}_\infty \mathrm{d}\mathbf{Q} = \mathbf{P}(A),$$

即 $\mathrm{d}\mathbf{P}/\mathrm{d}\mathbf{Q} = \mathfrak{z}_\infty$.

从而, 这就证明了自然期待的结果: *假如测度 $\mathbf{P}$ 和 $\mathbf{Q}$ 定义在 $\mathscr{F} = \sigma\left(\bigcup_n \mathscr{F}_n\right)$ 上, 而 $\mathbf{P}_n$ 和 $\mathbf{Q}_n$ 是这些测度到 $\mathscr{F}_n$ 上的收缩, 则关于测度 $\mathbf{Q}$ 几乎处处和在 $L^1(\Omega, \mathscr{F}, \mathbf{Q})$ 收敛的意义上, 有*

$$\lim_n \frac{\mathrm{d}\mathbf{P}_n}{\mathrm{d}\mathbf{Q}_n} = \frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}}.$$

类似地, 有

$$\lim_n \frac{\mathrm{d}\widetilde{\mathbf{P}}_n}{\mathrm{d}\mathbf{Q}_n} = \frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{Q}}.$$

在我们所讨论的特别情形下, 即当 $\widetilde{\mathbf{P}}_n \ll \mathbf{P}_n, n \geqslant 1$, 时, 不难证明, $(\mathbf{Q}\text{–a.c.})$ 有

$$z_n = \frac{\widetilde{\mathfrak{z}}}{\mathfrak{z}}, \tag{6}$$

这时, 由于

$$\mathbf{Q}\{\mathfrak{z}_n = 0, \widetilde{\mathfrak{z}}_n = 0\} \leqslant \frac{1}{2}[\mathbf{P}\{\mathfrak{z}_n = 0\} + \widetilde{\mathbf{P}}\{\widetilde{\mathfrak{z}}_n = 0\}] = 0,$$

因此, 可见在 (6) 式中 ($\mathbf{Q}$−a.c.) 不会出现形如 0/0 不定式的情形.

形如 2/0 的记号, 像通常一样认为等于 $+\infty$. 需要指出, 因为 $(z_n, \mathscr{F}_n)$ 是非负鞅, 所以由 §2 的 (5) 式可见, 如果 $z_\tau = 0$, 则对于一切 $n \geqslant \tau$ ($\mathbf{Q}$−a.c.), $z_n = 0$. 当然, 同样对于 $(\widetilde{z}_n, \mathscr{F}_n)$ 也是一样. 由此可见, 点 0 和 $+\infty$ 也是序列 $(z_n)_{n\geqslant 1}$ 的 "吸收状态".

由于 (5) 式和 (6) 式可见, $\mathbf{Q}$−a.c. 存在极限

$$z_\infty \equiv \lim_n z_n = \frac{\lim_n \widetilde{\mathfrak{z}}_n}{\lim_n \mathfrak{z}_n} = \frac{\widetilde{\mathfrak{z}}_\infty}{\mathfrak{z}_\infty}. \tag{7}$$

因为

$$\mathbf{P}\{\mathfrak{z}_\infty = 0\} = \int_{\{\mathfrak{z}_\infty = 0\}} \mathfrak{z}_\infty \mathrm{d}\mathbf{Q} = 0,$$

所以 $\mathbf{P}\{z_\infty = \infty\} = 0$, 因而定理的命题 a) 得证.

为证明 (3) 式, 利用 (4) 式的一般分解. 对于现在所考虑的情形, 由已经证明的结果:

$$\mathfrak{z} = \frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}} = \mathfrak{z}_\infty, \quad \widetilde{\mathfrak{z}} = \frac{\mathrm{d}\widetilde{\mathbf{P}}}{\mathrm{d}\mathbf{Q}} = \widetilde{\mathfrak{z}}_\infty \quad (\mathbf{Q} - \text{a.c.}),$$

知由 (4) 式, 得

$$\widetilde{\mathbf{P}}(A) = \int_A \frac{\widetilde{\mathfrak{z}}_\infty}{\mathfrak{z}_\infty} \mathrm{d}\mathbf{P} + \widetilde{\mathbf{P}}(A \bigcap \{\mathfrak{z}_\infty = 0\});$$

由 (7) 式, 以及由于 $\widetilde{\mathbf{P}}\{\widetilde{\mathfrak{z}}_\infty = 0\} = 0$, 得所要证明的 (3) 式. 最后注意到, 由于 $\mathbf{P}\{z_\infty < \infty\} = 1$, 可见对于 $A \in \mathscr{F}$, 则测度

$$\mathbf{P}(A) \equiv \mathbf{P}(A \bigcap \{z_\infty < \infty\}) \quad 和 \quad \widetilde{\mathbf{P}}(A \bigcap \{z_\infty = \infty\})$$

是奇异的. □

由勒贝格分解 (3), 可得如下关于局部绝对连续概率测度的, 绝对连续性和奇异性重要的准则.

定理 2 设 $\widetilde{\mathbf{P}} \overset{\text{loc}}{\ll} \mathbf{P}$, 即 $\widetilde{\mathbf{P}}_n \ll \mathbf{P}_n, n \geqslant 1$. 那么,

$$\widetilde{\mathbf{P}} \ll \mathbf{P} \Leftrightarrow \mathbf{E} z_\infty = 1 \Leftrightarrow \widetilde{\mathbf{P}}\{z_\infty < \infty\} = 1, \tag{8}$$

$$\widetilde{\mathbf{P}} \perp \mathbf{P} \Leftrightarrow \mathbf{E} z_\infty = 0 \Leftrightarrow \widetilde{\mathbf{P}}\{z_\infty = \infty\} = 1, \tag{9}$$

其中 $\mathbf{E}$ 表示对测度 $\mathbf{P}$ 的平均 (数学期望).

证明 在 (3) 式中设 $A = \Omega$, 得

$$\mathbf{E} z_\infty = 1 \Leftrightarrow \widetilde{\mathbf{P}}\{z_\infty = \infty\} = 0, \tag{10}$$

$$\mathbf{E} z_\infty = 0 \Leftrightarrow \widetilde{\mathbf{P}}\{z_\infty = \infty\} = 1. \tag{11}$$

如果 $\widetilde{\mathbf{P}}\{z_\infty=\infty\}=0$, 则仍然由 (3) 式可见 $\widetilde{\mathbf{P}}\ll\mathbf{P}$.

相反, 假设 $\widetilde{\mathbf{P}}\ll\mathbf{P}$. 那么, 由于 $\mathbf{P}\{z_\infty=\infty\}=0$, 可见 $\widetilde{\mathbf{P}}\{z_\infty=\infty\}=0$.

其次, 假如 $\widetilde{\mathbf{P}}\perp\mathbf{P}$, 则存在集合 $B\in\mathscr{F}$, 使 $\widetilde{\mathbf{P}}(B)=1$ 和 $\mathbf{P}(B)=0$. 那么, 由 (3) 式可见 $\widetilde{\mathbf{P}}(B\bigcap\{z_\infty=\infty\})=1$, 因此 $\widetilde{\mathbf{P}}\{z_\infty=\infty\}=1$. 相反, 假如 $\widetilde{\mathbf{P}}\{z_\infty=\infty\}=1$, 则因为 $\widetilde{\mathbf{P}}\{z_\infty=\infty\}=0$, 所以性质 $\widetilde{\mathbf{P}}\perp\mathbf{P}$ 显然. □

2. 应用绝对连续性和奇异性准则的例 由定理 2 知, 绝对连续性和奇异性准则, 或者可以用测度 $\mathbf{P}$ 的术语表示 (并验证等式 $\mathbf{E}z_\infty=1$ 或 $\mathbf{E}z_\infty=0$), 或者可以通过测度 $\widetilde{\mathbf{P}}$ 的术语表示 (并验证等式 $\widetilde{\mathbf{P}}\{z_\infty<\infty\}=1$ 或 $\widetilde{\mathbf{P}}\{z_\infty=\infty\}=1$).

由第二章 §6 定理 5 知, 条件 $\mathbf{E}z_\infty=1$ 等价于条件 "随机变量族 $\{z_n\}_{n\geqslant1}$ (关于测度 $\mathbf{P}$) 一致可积". 这一情况, 使得可以给出绝对连续性 $\widetilde{\mathbf{P}}\ll\mathbf{P}$ 简单的充分条件. 例如, 若

$$\sup_n \mathbf{E}[z_n\ln^+ z_n]<\infty \tag{12}$$

或如果

$$\sup_n \mathbf{E}[z_n^{1+\varepsilon}]<\infty,\quad \varepsilon>0, \tag{13}$$

则由第二章 §6 引理 3 知, 随机变量族 $\{z_n\}_{n\geqslant1}$ 一致可积, 因此 $\widetilde{\mathbf{P}}\ll\mathbf{P}$.

在许多情形下, 更倾向于运用通过测度 $\widetilde{\mathbf{P}}$ 的术语表示的准则, 因为这时可以归结为 "尾部" 事件 $\{z_\infty<\infty\}$ 之 $\widetilde{\mathbf{P}}-$ 概率的研究, 而这里可以利用 "0–1" 律类型的论点. 作为例子, 我们现在讨论如何由定理 2 导出角谷择一性.

设 $(\Omega,\mathscr{F},\mathbf{P})$ 是某一概率空间, $(\mathbb{R}^\infty,\mathscr{B}_\infty)$ 是数列 $x=(x_1,x_2,\cdots)$ 的可测空间, 其中 $\mathscr{B}_\infty=\mathscr{B}(\mathbb{R}^\infty)$, 而 $\mathscr{B}_n=\sigma(x_1,\cdots,x_n)$. 假设 $\xi=(\xi_1,\xi_2,\cdots)$ 和 $\widetilde{\xi}=(\widetilde{\xi}_1,\widetilde{\xi}_2,\cdots)$ 是两个独立随机变量序列.

记 P 和 $\widetilde{P}$ 相应为随机变量 ξ 和 $\widetilde{\xi}$ 在 $(\mathbb{R}^\infty,\mathscr{B}_\infty)$ 上的概率分布, 即

$$P(B)=\mathbf{P}\{\xi\in B\},\quad \widetilde{P}(B)=\mathbf{P}\{\widetilde{\xi}\in B\},\quad B\in\mathscr{B}_\infty.$$

亦假设

$$P_n=P|\mathscr{B}_n,\widetilde{P}_n=\widetilde{P}|\mathscr{B}_n$$

是 P 和 $\widetilde{P}$ 测度在 $\mathscr{B}_n$ 上的收缩, 且对于 $A\in\mathscr{B}(\mathbb{R}^1)$,

$$P_{\xi_n}(A)=\mathbf{P}\{\xi_n\in A\},$$
$$P_{\widetilde{\xi}_n}(A)=\mathbf{P}\{\widetilde{\xi}_n\in B\}.$$

定理 3 (角谷 (S. Kakutani) 择一性) 假设 $\xi=(\xi_1,\xi_2,\cdots)$ 和 $\widetilde{\xi}=(\widetilde{\xi}_1,\widetilde{\xi}_2,\cdots)$ 是两个独立随机变量序列, 且满足条件

$$P_{\widetilde{\xi}_n}\ll P_{\xi_n},\quad n\geqslant1. \tag{14}$$

那么, 要么 $\widetilde{P}\ll P$, 要么 $\widetilde{P}\perp P$.

证明 显然, 条件 (14) 等价于条件: $\widetilde{P}_n \ll P_n, n \geqslant 1$, 即 $\widetilde{P} \overset{\text{loc}}{\ll} P$.

$$z_n = \frac{\mathrm{d}\widetilde{P}_n}{\mathrm{d}P_n} = q_1(x_1)\cdots q_n(x_n),$$

其中

$$q_i(x_i) = \frac{\mathrm{d}P_{\widetilde{\xi}_i}}{\mathrm{d}P_{\xi_i}}(x_i). \tag{15}$$

从而

$$\{x : z_\infty < \infty\} = \{x : \ln z_\infty < \infty\} = \left\{x : \sum_{i=1}^{\infty} \ln q_i(x_i) < \infty\right\}.$$

事件

$$\left\{x : \sum_{i=1}^{\infty} \ln q_i(x_i) < \infty\right\}$$

是 "尾部的". 因此, 由柯尔莫戈洛夫 "0–1" 律 (第四章 §1 定理 1) 知, 概率 $\widetilde{P}\{x : z_\infty < \infty\}$ 只有 "0 或 1" 两个可能值. 于是, 根据定理 2, 要么 $\widetilde{P} \ll P$, 要么 $\widetilde{P} \perp P$.□

3. 测度的绝对连续性和奇异性与 "可预测性" 下面的定理用 "可预测的" 语言, 给出绝对连续性和奇异性的准则.

定理 4 设 $\widetilde{\mathbf{P}} \overset{\text{loc}}{\ll} \mathbf{P}$,

$$\alpha_n = z_n z_{n-1}^{\oplus}, \quad n \geqslant 1,$$

其中 $z_0 = 1$. 那么 ($\mathscr{F}_0 = \{\varnothing, \Omega\}$),

$$\widetilde{\mathbf{P}} \ll \mathbf{P} \Leftrightarrow \widetilde{\mathbf{P}}\left\{\sum_{n=1}^{\infty}[1 - \mathbf{E}(\sqrt{\alpha_n}|\mathscr{F}_{n-1})] < \infty\right\} = 1, \tag{16}$$

$$\widetilde{\mathbf{P}} \perp \mathbf{P} \Leftrightarrow \widetilde{\mathbf{P}}\left\{\sum_{n=1}^{\infty}[1 - \mathbf{E}(\sqrt{\alpha_n}|\mathscr{F}_{n-1})] = \infty\right\} = 1. \tag{17}$$

证明 因为

$$\widetilde{\mathbf{P}}_n\{z_n = 0\} = \int_{\{z_n=0\}} z_n \mathrm{d}\mathbf{P} = 0,$$

所以 ($\widetilde{\mathbf{P}}$ – a.c.)

$$z_n = \prod_{k=1}^{n} \alpha_k = \exp\left\{\sum_{k=1}^{n} \ln \alpha_k\right\}. \tag{18}$$

在 (3) 式中设 $A = \{z_\infty = 0\}$, 得 $\widetilde{\mathbf{P}}\{z_\infty = 0\} = 0$. 因此, 由 (18) 式, ($\mathbf{P}$ – a.c.) 有

$$\begin{aligned}\{z_\infty < \infty\} &= \{0 < z_\infty < \infty\} = \{0 < \lim z_n < \infty\} \\ &= \left\{-\infty < \lim \sum_{k=1}^{n} \ln \alpha_k < \infty\right\}.\end{aligned} \tag{19}$$

引进函数

$$u(x)=\begin{cases}x, & |x|\leqslant 1,\\ \text{sign }x, & |x|>1.\end{cases}$$

那么,

$$\left\{-\infty<\lim\sum_{k=1}^{n}\ln\alpha_k<\infty\right\}=\left\{-\infty<\lim\sum_{k=1}^{n}u(\ln\alpha_k)<\infty\right\}. \tag{20}$$

以 $\widetilde{\mathbf{E}}$ 表示对测度 $\widetilde{\mathbf{P}}$ 的数学期望, 而 η 是 $\mathscr{F}_n-$ 可测可积随机变量. 由条件数学期望的性质知 (练习题 4),

$$z_{n-1}\widetilde{\mathbf{E}}(\eta|\mathscr{F}_{n-1})=\mathbf{E}(\eta z_n|\mathscr{F}_{n-1})(\mathbf{P}-\text{ 和 }\widetilde{\mathbf{P}}-\text{a.c.}), \tag{21}$$

$$\widetilde{\mathbf{E}}(\eta|\mathscr{F}_{n-1})=z_{n-1}^{\oplus}\mathbf{E}(\eta z_n|\mathscr{F}_{n-1})\ (\widetilde{\mathbf{P}}-\text{a.c.}). \tag{22}$$

注意到, $\alpha_n=z_{n-1}^{\oplus}z_n$, 由 (22) 式得 "条件数学期望的换算公式" 的如下重要变式 (第二章 §7 的 (44) 式):

$$\widetilde{\mathbf{E}}(\eta|\mathscr{F}_{n-1})=\mathbf{E}(\alpha_n\eta|\mathscr{F}_{n-1})\quad(\widetilde{\mathbf{P}}-\text{a.c.}). \tag{23}$$

特别, 由 (23) 式可得

$$\mathbf{E}(\alpha_n|\mathscr{F}_{n-1})=1\quad(\widetilde{\mathbf{P}}-\text{a.c.}). \tag{24}$$

由 (23) 式, 有

$$\widetilde{\mathbf{E}}[u(\ln\alpha_n)|\mathscr{F}_{n-1}]=\mathbf{E}[\alpha_n u(\ln\alpha_n)|\mathscr{F}_{n-1}]\quad(\widetilde{\mathbf{P}}-\text{a.c.}).$$

因为对于一切 $x\geqslant 0$, 有 $xu(\ln x)\geqslant x-1$, 故由 (24) 式, 有

$$\widetilde{\mathbf{E}}[u(\ln\alpha_n)|\mathscr{F}_{n-1}]\geqslant 0\ (\widetilde{\mathbf{P}}-\text{a.c.}).$$

由此可见, 随机序列 $X=(X_n,\mathscr{F}_n)$, 其中

$$X_n=\sum_{k=1}^{n}u(\ln\alpha_k),$$

关于测度 $\widetilde{\mathbf{P}}$ 是下鞅, 且 $|\Delta X_n|=|u(\ln\alpha_k)|\leqslant 1$.

那么, 根据 §5 的定理 5 ($\widetilde{\mathbf{P}}-\text{a.c.}$) 有

$$\begin{aligned}&\left\{-\infty<\lim\sum_{k=1}^{n}u(\ln\alpha_k)<\infty\right\}\\ &=\left\{\sum_{k=1}^{\infty}\widetilde{\mathbf{E}}[u(\ln\alpha_k)+u^2(\ln\alpha_k)|\mathscr{F}_{k-1}]<\infty\right\}.\end{aligned} \tag{25}$$

这样, 由 (19), (20), (22) 和 (25) 式, ($\widetilde{\mathbf{P}}$ − a.c.) 有

$$\{z_\infty < \infty\} = \left\{\sum_{k=1}^{\infty} \widetilde{\mathbf{E}}[u(\ln \alpha_k) + u^2(\ln \alpha_k)|\mathscr{F}_{k-1}] < \infty\right\}$$
$$= \left\{\sum_{k=1}^{\infty} \mathbf{E}[\alpha_k u(\ln \alpha_k) + \alpha_k u^2(\ln \alpha_k)|\mathscr{F}_{k-1}] < \infty\right\},$$

从而, 由定理 2, 有

$$\widetilde{\mathbf{P}} \ll \mathbf{P} \Leftrightarrow \widetilde{\mathbf{P}}\left\{\sum_{k=1}^{\infty} \mathbf{E}[\alpha_k u(\ln \alpha_k) + \alpha_k u^2(\ln \alpha_k)|\mathscr{F}_{k-1}] < \infty\right\} = 1, \tag{26}$$

$$\widetilde{\mathbf{P}} \perp \mathbf{P} \Leftrightarrow \widetilde{\mathbf{P}}\left\{\sum_{k=1}^{\infty} \mathbf{E}[\alpha_k u(\ln \alpha_k) + \alpha_k u^2(\ln \alpha_k)|\mathscr{F}_{k-1}] = \infty\right\} = 1. \tag{27}$$

现在注意到, 由 (24) 式, 有

$$\widetilde{\mathbf{E}}[(1-\sqrt{\alpha_k})^2|\mathscr{F}_{k-1}] = 2\widetilde{\mathbf{E}}[1-\sqrt{\alpha_k}|\mathscr{F}_{k-1}] \ (\widetilde{\mathbf{P}} - \text{a.c.}),$$

且对于一切 $x \geqslant 0$, 存在常数 A 和 $B(0 < A < B < \infty)$, 使

$$A(1-\sqrt{x})^2 \leqslant xu(\ln x) + xu^2(\ln x) + 1 - x \leqslant B(1-\sqrt{x})^2. \tag{28}$$

于是, 由 (26), (27) 和 (24), (28) 式, 得结论 (16) 和 (17)式. □

系 1 假如对于 $\widetilde{\mathbf{P}} \overset{\text{loc}}{\ll} \mathbf{P}$, 及任意 $n \geqslant 1$, $\sigma-$ 代数 $\sigma(\alpha_n)$ 和 $\mathscr{F}_{n-1}$ (关于测度 $\mathbf{P}$ (或 $\widetilde{\mathbf{P}}$)) 独立, 则 $\widetilde{\mathbf{P}} \ll \mathbf{P}$ 或 $\widetilde{\mathbf{P}} \perp \mathbf{P}$ 两种情形必有一个并且只有一个成立. 这时

$$\widetilde{\mathbf{P}} \ll \mathbf{P} \Leftrightarrow \sum_{m=1}^{\infty} [1 - \widetilde{\mathbf{E}}\sqrt{\alpha_m}] < \infty,$$
$$\widetilde{\mathbf{P}} \perp \mathbf{P} \Leftrightarrow \sum_{m=1}^{\infty} [1 - \widetilde{\mathbf{E}}\sqrt{\alpha_m}] = \infty.$$

特别, 对于角谷的情形 (见定理 3), $\alpha_n = q_n$ 而

$$\widetilde{\mathbf{P}} \ll \mathbf{P} \Leftrightarrow \sum_{m=1}^{\infty} [1 - \mathbf{E}\sqrt{q_m(x_m)}] < \infty.$$
$$\mathbf{P} \perp \mathbf{P} \Leftrightarrow \sum_{m=1}^{\infty} [1 - \mathbf{E}\sqrt{q_m(x_m)}] = \infty.$$

系 2 设 $\widetilde{\mathbf{P}} \overset{\text{loc}}{\ll} \mathbf{P}$, 则

$$\widetilde{\mathbf{P}}\left\{\sum_{m=1}^{\infty} \mathbf{E}(\alpha_m \ln \alpha_m|\mathscr{F}_{m-1}) < \infty\right\} = 1 \Rightarrow \widetilde{\mathbf{P}} \ll \mathbf{P}.$$

为证明, 只需注意到, 对于任意 $x \geqslant 0$, 有

$$x \ln x + \frac{3}{2}(1-x) \geqslant 1 - x^{1/2}, \tag{29}$$

并利用 (16) 和 (17) 式.

系 3 因为 ($\widetilde{\mathbf{P}}$ − a.c.) 非负项级数

$$\sum_{n=1}^{\infty}[1-\mathbf{E}(\sqrt{\alpha_n}|\mathscr{F}_{n-1})] \quad 与 \quad \sum_{n=1}^{\infty}|\ln \mathbf{E}(\sqrt{\alpha_n}|\mathscr{F}_{n-1})|$$

同时收敛或发散, 则定理 4 的论断 (16) 和 (17) 式可以写成如下的形式:

$$\widetilde{\mathbf{P}} \ll \mathbf{P} \Leftrightarrow \widetilde{\mathbf{P}}\left\{\sum_{n=1}^{\infty}|\ln \mathbf{E}(\sqrt{\alpha_n}|\mathscr{F}_{n-1})| < \infty\right\} = 1, \tag{30}$$

$$\widetilde{\mathbf{P}} \perp \mathbf{P} \Leftrightarrow \widetilde{\mathbf{P}}\left\{\sum_{n=1}^{\infty}|\ln \mathbf{E}(\sqrt{\alpha_n}|\mathscr{F}_{n-1})| = \infty\right\} = 1. \tag{31}$$

系 4 设存在常数 $0 \leqslant A < 1$ 和 $B \geqslant 0$, 使

$$\mathbf{P}\{1-A \leqslant \alpha_n \leqslant 1+B\} = 1, \quad n \geqslant 1.$$

那么, 如果 $\widetilde{\mathbf{P}} \overset{\text{loc}}{\ll} \mathbf{P}$, 则

$$\widetilde{\mathbf{P}} \ll \mathbf{P} \Leftrightarrow \widetilde{\mathbf{P}}\left\{\sum_{n=1}^{\infty}\mathbf{E}[(1-\alpha_n)^2|\mathscr{F}_{n-1}] < \infty\right\} = 1,$$

$$\widetilde{\mathbf{P}} \perp \mathbf{P} \Leftrightarrow \widetilde{\mathbf{P}}\left\{\sum_{n=1}^{\infty}\mathbf{E}[(1-\alpha_n)^2|\mathscr{F}_{n-1}] = \infty\right\} = 1.$$

为证明, 只需注意到, 对于 $x \in [1-A, 1+B], 0 \leqslant A < 1$ 和 $B \geqslant 0$, 存在常数 c 和 $C(0 < c < C < \infty)$, 使

$$c(1-x)^2 \leqslant (1-\sqrt{x})^2 \leqslant C(1-x)^2. \tag{32}$$

4. 哈伊克 – 费里德曼择一性 假设 $\xi = (\xi_1, \xi_2, \cdots)$ 和 $\widetilde{\xi} = (\widetilde{\xi}_1, \widetilde{\xi}_2, \cdots)$ 是两个高斯随机变量序列, 且 (在第 2 小节的记号下) $\widetilde{P}_n \sim P_n, n \geqslant 1$. 我们说明, 对于这样的序列, 如何由上面得到的 “可预测” 准则, 得出 “哈伊克 – 费里德曼 (J. Hajek–H. M. Friedman) 择一性”: $\widetilde{P} \sim P$ 或 $\widetilde{P} \perp P$ 二者必居其一.

根据正态相关定理 (第二章 §13 定理 2), 条件数学期望 $\mathbf{E}(x_n|\mathscr{B}_{n-1})$ 和 $\widetilde{\mathbf{E}}(x_n|\mathscr{B}_{n-1})$ 是 $x_1, \cdots, x_{n-1}$ 的线性函数, 其中 $\mathbf{E}$ 和 $\widetilde{\mathbf{E}}$ 相应为对测度 P 和 $\widetilde{P}$ 求数学期望. 以 $a_{n-1}(x)$ 和 $\widetilde{a}_{n-1}(x)$ 分别表示这些 (线性) 函数 ($a_0(x) = a_0$ 和 $\widetilde{a}_0(x) = \widetilde{a}_0$ 是常数). 记

$$b_{n-1} = (\mathbf{E}[x_n - a_{n-1}(x)]^2)^{1/2},$$
$$\widetilde{b}_{n-1} = (\widetilde{\mathbf{E}}[x_n - \widetilde{a}_{n-1}(x)]^2)^{1/2}.$$

根据上面提到的正态相关定理, 存在均值为 0、方差为 1 的, 高斯随机变量序列 $\varepsilon=(\varepsilon_1,\varepsilon_2,\cdots)$ 和 $\widetilde{\varepsilon}=(\widetilde{\varepsilon}_1,\widetilde{\varepsilon}_2,\cdots)$, 使得 $(\mathbf{P}-\text{a.c.})$

$$\begin{aligned}\xi_n &= a_{n-1}(\xi)+b_{n-1}\varepsilon_n,\\ \widetilde{\xi}_n &= \widetilde{a}_{n-1}(\widetilde{\xi})+\widetilde{b}_{n-1}\widetilde{\varepsilon}_n.\end{aligned} \tag{33}$$

注意, 在 $b_{n-1}=0(\widetilde{b}_{n-1}=0)$ 情形下, 为构造随机变量 $\varepsilon_n(\widetilde{\varepsilon}_n)$, 一般不得不扩充概率空间. 不过, 假如 $b_{n-1}=0$, 则随机向量 $(x_1,\cdots,x_n)$ 的分布 $(P-\text{a.c.})$ 集中在线性流形 $x_n=a_{n-1}(x)$ 上. 因为根据假设 $\widetilde{P}_n\sim P_n$, 所以 $\widetilde{b}_{n-1}=0, a_{n-1}(x)=\widetilde{a}_{n-1}(x)$ 和 $\alpha_n(x)=1$ $(P-$ 及 $\widetilde{P}-\text{a.c.})$. 因此, 不失普遍性可以认为, 对于一切 $n\geqslant 1$, 有 $b_n^2>0,\widetilde{b}_n^2>0$, 因为在相反的情形下, 和式

$$\sum_{n=1}^{\infty}[1-\mathbf{E}(\sqrt{\alpha_n}|\mathscr{B}_{n-1})]$$

中相应项的 "贡献" 等于 0 (见 (16) 式和 (17) 式).

利用高斯性假设, 由 (33) 式可见, 对于 $n\geqslant 1$, 有

$$\alpha_n=d_{n-1}^{-1}\exp\left\{-\frac{[x_n-a_{n-1}(x)]^2}{2b_{n-1}^2}+\frac{[x_n-\widetilde{a}_{n-1}(x)]^2}{2\widetilde{b}_{n-1}^2}\right\}, \tag{34}$$

其中 $d_n=|b_n/\widetilde{b}_n|$, 而

$$\begin{aligned}a_0=\mathbf{E}\xi_1,\quad \widetilde{a}_0=\mathbf{E}\widetilde{\xi}_1,\\ b_0^2=\mathbf{D}\xi_1,\quad \widetilde{b}_0^2=\mathbf{D}\widetilde{\xi}_1.\end{aligned}$$

由 (34), 有

$$\ln\mathbf{E}(\alpha_n^{1/2}|\mathscr{B}_{n-1})=\frac{1}{2}\ln\frac{2d_{n-1}}{1+d_{n-1}^2}-\frac{d_{n-1}^2}{1+d_{n-1}^2}\left(\frac{a_{n-1}(x)-\widetilde{a}_{n-1}(x)}{b_{n-1}}\right)^2.$$

由于

$$\ln\frac{2d_{n-1}}{1+d_{n-1}^2}\leqslant 0$$

则 (30) 式有如下形式

$$\widetilde{P}\ll P\Leftrightarrow\widetilde{P}\left\{\sum_{n=1}^{\infty}\left[\frac{1}{2}\ln\frac{1+d_{n-1}^2}{2d_{n-1}}+\frac{d_{n-1}^2}{1+d_{n-1}^2}\left(\frac{a_{n-1}(x)-\widetilde{a}_{n-1}(x)}{b_{n-1}}\right)^2\right]<\infty\right\}=1. \tag{35}$$

由于级数

$$\sum_{n=1}^{\infty}\ln\frac{1+d_{n-1}^2}{2d_{n-1}}\quad \text{和}\quad \sum_{n=1}^{\infty}(d_{n-1}^2-1)^2$$

或者同时收敛, 或者同时发散, 则由 (35) 式可见

$$\widetilde{P} \ll P \Leftrightarrow \widetilde{P}\left\{\sum_{n=1}^{\infty}\left[\left(\frac{\Delta_n(x)}{b_n}\right)^2+\left(\frac{\widetilde{b}_n^2}{b_n^2}-1\right)^2\right]<\infty\right\}=1, \tag{36}$$

其中 $\Delta_n(x)=a_n(x)-\widetilde{a}_n(x)$.

因为 $a_n(x)$ 和 $\widetilde{a}_n(x)$ 是线性函数, 所以随机变量序列 $\{\Delta_n(x)/b_n\}_{n\geqslant 0}$ (既关于测度 $\widetilde{P}$, 也关于测度 P) 构成高斯系统. 由下面将要证明的引理, 可见

$$\widetilde{P}\left\{\sum\left(\frac{\Delta_n(x)}{b_n}\right)^2<\infty\right\}=1 \Leftrightarrow \sum \widetilde{\mathbf{E}}\left(\frac{\Delta_n(x)}{b_n}\right)^2<\infty. \tag{37}$$

从而, 由 (36) 式, 可见

$$\widetilde{P} \ll P \Leftrightarrow \sum_{n=0}^{\infty}\left[\widetilde{\mathbf{E}}\left(\frac{\Delta_n(x)}{b_n}\right)^2+\left(\frac{\widetilde{b}_n^2}{b_n^2}-1\right)^2\right]<\infty.$$

类似地, 可得

$$\widetilde{P} \perp P \Leftrightarrow \widetilde{P}\left\{\sum_{n=1}^{\infty}\left[\left(\frac{\Delta_n(x)}{b_n}\right)^2+\left(\frac{\widetilde{b}_n^2}{b_n^2}-1\right)^2\right]<\infty\right\}=0$$
$$\Leftrightarrow \sum_{n=0}^{\infty}\left[\widetilde{\mathbf{E}}\left(\frac{\Delta_n(x)}{b_n}\right)^2+\left(\frac{\widetilde{b}_n^2}{b_n^2}-1\right)^2\right]=\infty.$$

由此可见, 如果测度 P 与 $\widetilde{P}$ 非奇异, 则 $\widetilde{P} \ll P$. 因为根据假设 $\widetilde{P}_n \sim P, n \geqslant 1$, 所以由对称性知 $P \ll \widetilde{P}$. 于是有下面的定理.

定理 5 (哈伊克 – 费里德曼择一性) 设 $\xi=(\xi_1,\xi_2,\cdots)$ 和 $\widetilde{\xi}=(\widetilde{\xi}_1,\widetilde{\xi}_2,\cdots)$ 是两个高斯随机变量序列, 且其有限维分布等价: $\widetilde{P}_n \sim P_n, n \geqslant 1$. 那么, $\widetilde{P} \sim P$ 或 $\widetilde{P} \perp P$ 二者必居其一. 这时

$$\begin{aligned}
\widetilde{P} \sim P &\Leftrightarrow \sum_{n=0}^{\infty}\left[\widetilde{\mathbf{E}}\left(\frac{\Delta_n(x)}{b_n}\right)^2+\left(\frac{\widetilde{b}_n^2}{b_n^2}-1\right)^2\right]<\infty,\\
\widetilde{P} \perp P &\Leftrightarrow \sum_{n=0}^{\infty}\left[\widetilde{\mathbf{E}}\left(\frac{\Delta_n(x)}{b_n}\right)^2+\left(\frac{\widetilde{b}_n^2}{b_n^2}-1\right)^2\right]=\infty.
\end{aligned} \tag{38}$$

引理 设 $\beta=(\beta_n)_{n\geqslant 1}$ 是定义在 $(\Omega,\mathscr{F},\mathbf{P})$ 上的高斯序列. 那么,

$$\mathbf{P}\left\{\sum_{n=1}^{\infty}\beta_n^2<\infty\right\}>0 \Leftrightarrow \mathbf{P}\left\{\sum_{n=1}^{\infty}\beta_n^2<\infty\right\}=1 \Leftrightarrow \sum_{n=1}^{\infty}\mathbf{E}\beta_n^2<\infty. \tag{39}$$

证明 蕴涵关系 ($\Leftarrow$) 显然. 现在假设 $\mathbf{E}\beta_n^2 = 0, n \geqslant 1$, 并证明蕴涵关系 ($\Rightarrow$). 为此, 只需证明

$$\mathbf{E}\sum_{n=1}^{\infty}\beta_n^2 \leqslant \left[\mathbf{E}\exp\left(-\sum_{n=1}^{\infty}\beta_n^2\right)\right]^{-2}, \tag{40}$$

因为, 由 $\mathbf{P}\{\sum\beta_n^2 < \infty\} > 0$ 条件可见, (40) 式的右侧小于 ∞. 即 $\sum\limits_{n=1}^{\infty}\mathbf{E}\beta_n^2 < \infty$, 而由已证明的蕴涵关系, 可见

$$\mathbf{P}\left\{\sum_{n=1}^{\infty}\beta_n^2 < \infty\right\} = 1.$$

固定某一 $n \geqslant 1$, 则由第二章 §11 和 §13 知, 存在独立高斯随机变量 $\beta_{k,n}(k = 1, \cdots, r \leqslant n)$, 且 $\mathbf{E}\beta_{k,n} = 0$, 使得

$$\sum_{k=1}^{n}\beta_k^2 = \sum_{n=1}^{r}\beta_{k,n}^2.$$

如果记 $\mathbf{E}\beta_{k,n}^2 = \lambda_{k,n}$, 则易见

$$\mathbf{E}\sum_{k=1}^{r}\beta_{k,n}^2 = \sum_{k=1}^{r}\lambda_{k,n} \tag{41}$$

而

$$\mathbf{E}\exp\left(-\sum_{k=1}^{r}\beta_{k,n}^2\right) = \prod_{k=1}^{r}(1+2\lambda_{k,n})^{-1/2}. \tag{42}$$

比较 (41) 和 (42) 式右侧, 得

$$\mathbf{E}\sum_{k=1}^{n}\beta_k^2 = \mathbf{E}\sum_{k=1}^{r}\beta_{k,n}^2 \leqslant \left[\mathbf{E}\exp\left(-\sum_{k=1}^{r}\beta_{k,n}^2\right)\right]^{-2} = \left[\mathbf{E}\exp\left(-\sum_{k=1}^{n}\beta_k^2\right)\right]^{-2},$$

由此, 当 $n \to \infty$ 时经极限过渡, 得所要证明得不等式 (40).

现在假设 $\mathbf{E}\beta_n \neq 0$.

考虑与 $\beta = (\beta_n)_{n\geqslant 1}$ 独立并且具有同一分布的序列 $\widetilde{\beta} = (\widetilde{\beta}_n)_{n\geqslant 1}$ (必要时扩展原概率空间). 那么, 如果

$$\mathbf{P}\left\{\sum_{n=1}^{\infty}\beta_n < \infty\right\} > 0, \quad 则 \quad \mathbf{P}\left\{\sum_{n=1}^{\infty}(\beta_n - \widetilde{\beta}_n)^2 < \infty\right\} > 0,$$

且由已经证明的, 有

$$2\sum_{n=1}^{\infty}\mathbf{E}(\beta_n - \mathbf{E}\beta_n)^2 = \sum_{n=1}^{\infty}\mathbf{E}(\beta_n - \widetilde{\beta}_n)^2 < \infty.$$

因为

$$(\mathbf{E}\beta_n)^2 \leqslant 2\beta_n^2 + 2(\beta_n - \mathbf{E}\beta_n)^2, \quad 则 \quad \sum_{n=1}^{\infty}(\mathbf{E}\beta_n)^2 < \infty.$$

于是,

$$\sum_{n=1}^{\infty}\mathbf{E}\beta_n^2 = \sum_{n=1}^{\infty}(\mathbf{E}\beta_n)^2 = \sum_{n=1}^{\infty}\mathbf{E}(\beta_n - \mathbf{E}\beta_n)^2 < \infty. \qquad \square$$

5. 例 我们继续讨论 §5 第 3 小节的例. 假设 $\xi_0, \xi_1, \cdots$ 是独立高斯随机变量序列, $\mathbf{E}\xi_i = 0, \mathbf{D}\xi_i = V_i > 0$.

仍设

$$X_{n+1} = \theta X_n + \xi_{n+1}, \quad n \geqslant 0,$$

其中 $X_0 = \xi_0$, 而 $\theta(-\infty < \theta < \infty)$ 是待估计未知参数. 设 $\widehat{\theta}_n$ 是用最小二乘法得到的估计.

定理 6 *序列 $\widehat{\theta}_n(n \geqslant 1)$ 是强相合估计的, 充分和必要条件是:*

$$\sum_{n=0}^{\infty}\frac{V_n}{1+V_n} = \infty. \tag{43}$$

证明 *充分性.* 当未知参数的值为 θ 时, 以 $\mathbf{P}_\theta$ 表示 $(\mathbb{R}^\infty, \mathscr{B}^\infty)$ 上对应于序列 $(X_0, X_2, \cdots)$ 的概率分布; 以 $\mathbf{E}_\theta$ 表示对测度 $\mathbf{P}_\theta$ 的数学期望.

我们已经知道,

$$\widehat{\theta}_n = \theta + \frac{M_n}{\langle M\rangle_n},$$

其中

$$M_n = \sum_{k=0}^{n-1}\frac{X_k\xi_{k+1}}{V_{k+1}}, \quad \langle M\rangle_n = \sum_{k=0}^{n-1}\frac{X_k^2}{V_{k+1}}.$$

根据上一小节的引理

$$\mathbf{P}_\theta\{\langle M\rangle_\infty = \infty\} = 1 \Leftrightarrow \mathbf{E}_\theta\langle M\rangle_\infty = \infty.$$

从而 $\langle M\rangle_\infty = \infty$ ($\mathbf{P}_\theta$ – a.c.), 当且仅当

$$\sum_{k=1}^{\infty}\frac{\mathbf{E}_\theta X_k^2}{V_{k+1}} = \infty. \tag{44}$$

而且

$$\mathbf{E}_\theta X_k^2 = \sum_{i=0}^{k}\theta^{2i}V_{k-i},$$

$$\begin{aligned}\sum_{k=0}^{\infty}\frac{\mathbf{E}_\theta X_k^2}{V_{k+1}} &= \sum_{k=0}^{\infty}\frac{1}{V_{k+1}}\left(\sum_{i=0}^{k}\theta^{2i}V_{k-i}\right) = \sum_{i=k}^{\infty}\theta^{2k}\sum_{i=k}^{\infty}\frac{V_{i-k}}{V_{i+1}} \\ &= \sum_{i=0}^{\infty}\frac{V_i}{V_{i+1}} + \sum_{k=1}^{\infty}\theta^{2k}\left(\sum_{i=k}^{\infty}\frac{V_{i-k}}{V_{i+1}}\right).\end{aligned} \tag{45}$$

因此由 (43) 式得 (44) 式, 而根据定理 4, 对于每一个 θ, 估计 $\widehat{\theta}_n(n \geqslant 1)$ 序列为强相合的.

必要性 假设对于一切 $\theta, \mathbf{P}_\theta(\widehat{\theta}_n \to \theta) = 1$. 现在证明, 假如 $\theta_1 \neq \theta_2$, 则 $\mathbf{P}_{\theta_1}$ 和 $\mathbf{P}_{\theta_2}$ 是奇异的 $(\mathbf{P}_{\theta_1} \perp \mathbf{P}_{\theta_2})$. 事实上, 由于 $(X_0, X_1, \cdots)$ 是高斯序列, 则根据定理 5 测度 $\mathbf{P}_{\theta_1}$ 和 $\mathbf{P}_{\theta_2}$ 要么奇异要么等价. 然而测度 $\mathbf{P}_{\theta_1}$ 和 $\mathbf{P}_{\theta_2}$ 不可能等价, 因为, 假如 $\mathbf{P}_{\theta_1} \sim \mathbf{P}_{\theta_2}$, 但是由于 $\mathbf{P}_{\theta_1}(\widehat{\theta}_n \to \theta_1) \to 1$, 则 $\mathbf{P}_{\theta_2}(\widehat{\theta}_n \to \theta_1) = 1$. 而由于 $\mathbf{P}_{\theta_2}(\widehat{\theta}_n \to \theta_2) = 1$, 故应该 $\theta_1 = \theta_2$, 而这与假设 $\theta_1 \neq \theta_2$ 矛盾.

于是, 对于 $\theta_1 \neq \theta_2, \mathbf{P}_{\theta_1} \perp \mathbf{P}_{\theta_2}$.

根据 (38) 式, 对于 $\theta_1 \neq \theta_2$, 有

$$\mathbf{P}_{\theta_1} \perp \mathbf{P}_{\theta_2} \Leftrightarrow (\theta_1 - \theta_2)^2 \sum_{k=0}^{\infty} \mathbf{E}_{\theta_1}\left(\frac{X_k^2}{V_{k+1}}\right) = \infty.$$

设 $\theta_1 = 0$ 和 $\theta_2 \neq 0$, 则由 (45) 式得

$$\mathbf{P}_0 \perp \mathbf{P}_{\theta_2} \Leftrightarrow \sum_{i=0}^{\infty} \frac{V_i}{V_{i+1}} = \infty.$$

于是, 这就证明了必要性. □

6. 练习题

1. 证明等式 (6).

2. 设 $\widetilde{\mathbf{P}}_n \sim \mathbf{P}_n, n \geqslant 1$. 证明

$$\widetilde{\mathbf{P}} \sim \mathbf{P} \Leftrightarrow \widetilde{\mathbf{P}}\{z_\infty < \infty\} = \mathbf{P}\{z_\infty > 0\} = 1,$$
$$\widetilde{\mathbf{P}} \perp \mathbf{P} \Leftrightarrow \widetilde{\mathbf{P}}\{z_\infty = \infty\} = 1 \text{ 或 } \mathbf{P}\{z_\infty = 0\} = 1.$$

3. 设 $\widetilde{\mathbf{P}}_n \ll \mathbf{P}_n, n \geqslant 1$, 而 τ (关于 $(\mathscr{F}_n)$) 是停时, $\widetilde{\mathbf{P}}_\tau = \widetilde{\mathbf{P}}|\mathscr{F}_\tau$ 和 $\mathbf{P}_\tau = \mathbf{P}|\mathscr{F}_\tau$ 是测度 $\widetilde{\mathbf{P}}$ 和 $\mathbf{P}$ 到 $\sigma-$ 代数 $\mathscr{F}_\tau$ 的收缩. 证明 $\widetilde{\mathbf{P}}_\tau \ll \mathbf{P}_\tau$, 当且仅当 $\{\tau = \infty\} = \{z_\infty < \infty\}(\mathbf{P} - \text{a.c.})$. 特别, 如果 $\mathbf{P}\{\tau < \infty\} = 1$, 则 $\widetilde{\mathbf{P}}_\tau \ll \mathbf{P}_\tau$.

4. 证明 "换算公式" (21) 和 (22).

5. 证明不等式 (28), (29), (32).

6. 证明公式 (34).

7. 假设在第 2 小节中, 序列 $\xi = (\xi_1, \xi_2, \cdots)$ 和 $\widetilde{\xi} = (\widetilde{\xi}_1, \widetilde{\xi}_2, \cdots)$ 由独立同分布随机变量构成. 证明, 1) 如果 $\mathbf{P}_{\widetilde{\xi}_1} \ll \mathbf{P}_{\xi_1}$, 则 $\widetilde{\mathbf{P}} \ll \mathbf{P}$, 当且仅当测度 $\mathbf{P}_{\widetilde{\xi}_1} = \mathbf{P}_{\xi_1}$; 2) 如果 $\mathbf{P}_{\widetilde{\xi}_1} \ll \mathbf{P}_{\xi_1}$ 且 $\mathbf{P}_{\widetilde{\xi}_1} \neq \mathbf{P}_{\xi_1}$, 则 $\widetilde{\mathbf{P}} \perp \mathbf{P}$.

§7. 随机游动越出曲线边界的概率的渐近式

1. 概率 $\mathbf{P}\{\tau > n\}$ 的渐近式 设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量序列, $S_n =$

$\xi_1+\cdots+\xi_n$; $g=g(n)$, $n\geqslant 1, g(1)<0$, 是某一“边界”; 而

$$\tau=\inf\{n\geqslant 1: S_n<g(n)\}$$

是随机游动 $(S_n)_{n\geqslant 1}$ 首次处于边界 $g=g(n)$ 以下的时间. (像通常一样, 如果 $\{\cdot\}=\varnothing$, 则认为 $\tau=\infty$). 求时间 τ 的分布是相当困难的. 在这一节, 对于广泛的一类边界 $g=g(n)$, 并且 ξ_i 在服从正态分布的条件下, 求当 $n\to\infty$ 时概率 $\mathbf{P}\{\tau>n\}$ 的渐近式. 所用方法基于 “测度绝对连续替换” 的思想, 并且利用前面介绍的鞅和马尔可夫时间的一系列性质.

定理 1　设 $\xi_1,\xi_2,\cdots$ 是独立同分布的随机变量, $\xi_i\sim N(0,1)$. 假设边界 $g=g(n)$ 满足条件: $g(1)<0$, 而对于 $n\geqslant 2$,

$$0\leqslant \Delta g(n+1)\leqslant \Delta g(n), \tag{1}$$

其中 $\Delta g(n)=g(n)-g(n-1)$, 并且

$$\ln n=o\left(\sum_{k=2}^{n}[\Delta g(k)]^2\right),\quad n\to\infty. \tag{2}$$

那么,

$$\mathbf{P}\{\tau>n\}=\exp\left\{-\frac{1}{2}\sum_{k=2}^{n}[\Delta g(k)]^2(1+o(1))\right\},\quad n\to\infty. \tag{3}$$

在进行证明之前, 我们首先指出, 条件 (1) 和 (2) 成立, 例如, 若

$$g(n)=an^{\nu}+b,\quad \frac{1}{2}<\nu\leqslant 1,\quad a+b<0,\quad a>0,$$

或者 (对于充分大的 n)

$$g(n)=an^{\nu}L(n),\quad \frac{1}{2}\leqslant \nu\leqslant 1,$$

其中 $L(n)$ 是某一缓慢变化的函数 (例如, $L(n)=C(\ln n)^{\beta}$, 对于 $1/2<\nu<1$, β 是任意的, 而对于 $\nu=1/2, C>0,\beta>0$).

2. 定理 1 的证明　在证明定理 1 时用到下面两个辅助命题.

假设 $\xi_1,\xi_2,\cdots$ 是独立同分布的随机变量序列, $\xi_i\sim N(0,1)$. 引进记号: $\mathscr{F}_0=\{\varnothing,\Omega\},\mathscr{F}_n=\sigma\{\xi_1,\cdots,\xi_n\}$, 而设 $\alpha=(\alpha_n,\mathscr{F}_{n-1})$ 是可预测序列, 且 $\mathbf{P}\{|\alpha_n|\leqslant C\}=1, n\geqslant 1$, 其中 C 是某一常数. 建立序列 $z=(z_n,\mathscr{F}_n)$, 其中

$$z_n=\exp\left\{\sum_{k=1}^{n}\alpha_k\xi_k-\frac{1}{2}\sum_{k=1}^{n}\alpha_k^2\right\},\quad n\geqslant 1. \tag{4}$$

不难验证, 序列 (关于测度 $\mathbf{P}$) $z=(z_n,\mathscr{F}_n)$ 是鞅, 且 $\mathbf{E}z_n=1, n\geqslant 1$.

固定某个 $n\geqslant 1$, 并且在可测空间上 $(\Omega,\mathscr{F}_n)$引进概率测度 $\widetilde{\mathbf{P}}_n$:

$$\widetilde{\mathbf{P}}_n(A)=\mathbf{E}I(A)z_n,\quad A\in\mathscr{F}_n. \tag{5}$$

引理 1 (“士尔萨诺夫 (И. Гирсанов) 定理” 的离散变式) 随机变量 $\widetilde{\xi}_k = \xi_k - \alpha_k, 1 \leqslant k \leqslant n$, 关于测度 $\widetilde{\mathbf{P}}_n$, 独立且服从正态分布 $\xi_k \sim N(0,1)$.

证明 设 $\widetilde{\mathbf{E}}_n$ 表示对测度 $\widetilde{\mathbf{P}}_n$ 的数学期望, 则对于 $\lambda_k \in \mathbb{R}, 1 \leqslant k \leqslant n$, 有

$$
\begin{aligned}
&\widetilde{\mathbf{E}}_n \exp\left\{i\sum_{k=1}^{n}\lambda_k\widetilde{\xi}_k\right\} = \mathbf{E}\exp\left\{i\sum_{k=1}^{n}\lambda_k\widetilde{\xi}_k\right\} z_n \\
&= \mathbf{E}\left[\exp\left\{i\sum_{k=1}^{n}\lambda_k\widetilde{\xi}_k\right\} z_{n-1}\mathbf{E}\left\{\exp\left(i\lambda_n(\xi_n-\alpha_n)+\alpha_n\xi_n-\frac{\alpha_n^2}{2}\right)\Big|\mathscr{F}_{n-1}\right\}\right] \\
&= \mathbf{E}\left[\exp\left\{i\sum_{k=1}^{n}\lambda_k\widetilde{\xi}_k\right\} z_{n-1}\right]\exp\left\{-\frac{\lambda_n^2}{2}\right\} = \cdots = \exp\left\{-\frac{1}{2}\sum_{k=1}^{n}\lambda_k^2\right\}.
\end{aligned}
$$

于是, 由第二章 §12 定理 4 得所需要证明的结论. □

引理 2 设 $X = (X_n, \mathscr{F}_n)_{n\geqslant 1}$ 是均值为 0 的平方可积鞅, 而

$$\sigma = \inf\{n \geqslant 1 : X_n \leqslant -b\},$$

其中常数 $b > 0$. 假设

$$\mathbf{P}\{X_1 \leqslant -b\} > 0.$$

那么, 存在常数 $C > 0$, 使对于一切 $n \geqslant 1$,

$$\mathbf{P}\{\sigma > n\} \geqslant \frac{C}{\mathbf{E}X_n^2}. \tag{6}$$

证明 根据 §2 定理 1 的系 1, $\mathbf{E}X_{\sigma\wedge n} = 0$, 由此可见

$$-\mathbf{E}I(\sigma \leqslant n)X_\sigma = \mathbf{E}I(\sigma > n)X_n. \tag{7}$$

在集合 $\{\sigma \leqslant n\}$ 上, 有

$$-X_\sigma \geqslant b > 0.$$

因此, 对于 $n \geqslant 1$

$$-\mathbf{E}I(\sigma \leqslant n)X_\sigma \geqslant b\mathbf{P}\{\sigma \leqslant n\} \geqslant b\mathbf{P}\{\sigma = 1\} = b\mathbf{P}\{X_1 < -b\} > 0. \tag{8}$$

另一方面, 由于可西 – 布尼亚科夫斯基 (A. L. Cauchy – А.Я.Буняковский) 不等式, 可见

$$\mathbf{E}I(\sigma > n)X_n \leqslant [\mathbf{P}\{\sigma > n\}\cdot\mathbf{E}X_n^2]^{1/2}. \tag{9}$$

于是, 由 (9) 式及 (7) 和 (8) 式, 所要证明的不等式 (6), 其中 $C = [b\mathbf{P}\{X_1 < -b\}]^2$.

证明定理 1 只需证明:

$$\varliminf_n \ln\mathbf{P}\{\tau > n\}\Big/\sum_{k=2}^{n}[\Delta g(k)]^2 \geqslant -\frac{1}{2} \tag{10}$$

和

$$\overline{\lim_n} \ln \mathbf{P}\{\tau > n\} \Big/ \sum_{k=2}^{n} [\Delta g(k)]^2 \leqslant -\frac{1}{2}. \tag{11}$$

为此, 考虑 (非随机) 序列 $(\alpha_n)_{n\geqslant 1}$, 其中

$$\alpha_1 = 0, \quad \alpha_n = \Delta g(n), \quad n \geqslant 2,$$

而概率测度 $(\widetilde{\mathbf{P}}_n)_{n\geqslant 1}$ 决定于 (5) 式. 那么, 根据赫尔德 (O. L. Hölder) 不等式, 有

$$\widetilde{\mathbf{P}}_n\{\tau > n\} = \mathbf{E} I(\tau > n) z_n \leqslant (\mathbf{P}\{\tau > n\})^{1/q} (\mathbf{E} z_n^p)^{1/p}, \tag{12}$$

其中 $p > 1$ 和 $q = p/(p-1)$.

式 (12) 中最末尾一个因子的明显表达式为:

$$(\mathbf{E} z_n^p)^{1/p} = \exp\left\{\frac{p-1}{2} \sum_{k=2}^{n} [\Delta g(k)]^2\right\}. \tag{13}$$

现在估计 (12) 式左侧的概率 $\widetilde{\mathbf{P}}_n\{\tau > n\}$. 有

$$\widetilde{\mathbf{P}}_n\{\tau > n\} = \widetilde{\mathbf{P}}_n\{S_k \geqslant g(k), 1 \leqslant k \leqslant n\} = \widetilde{\mathbf{P}}_n\{\widetilde{S}_k \geqslant g(1), 1 \leqslant k \leqslant n\},$$

其中 $\widetilde{S}_k = \sum\limits_{i=1}^{n} \widetilde{\xi}_i$, $\widetilde{\xi}_i = \xi_i - \alpha_i$. 根据引理 1, 随机变量 $\widetilde{\xi}_1, \cdots, \widetilde{\xi}_n$, 关于测度 $\widetilde{\mathbf{P}}_n$ 独立且服从正态分布 $\widetilde{\xi}_k \sim N(0,1)$. 将引理 2 用于 $b = -g(1)$, $\mathbf{P} = \widetilde{\mathbf{P}}_n$, $X_n = \widetilde{S}_k$, 得

$$\widetilde{\mathbf{P}}_n\{\tau > n\} \geqslant \frac{C}{n}, \tag{14}$$

其中 C 是某一常数.

那么, 由 (12)~(14) 式, 可见对于任何 $p > 1$, 有

$$\mathbf{P}\{\tau > n\} \geqslant C_p \exp\left\{-\frac{p}{2} \sum_{k=2}^{n} [\Delta g(k)]^2 - \frac{p}{p-1} \ln n\right\}, \tag{15}$$

其中 C_p 是某一常数. 鉴于定理的条件和 $p > 1$ 的任意性, 由 (15) 式得的下侧估计 (10).

为得到 (11) 式的上侧估计, 首先注意到由于 $z_n > 0$ ($\mathbf{P}-$ 且 $\widetilde{\mathbf{P}}-$ a.c.), 则由 (5) 式可见

$$\mathbf{P}\{\tau > n\} = \widetilde{\mathbf{E}}_n I(\tau > n) z_n^{-1}, \tag{16}$$

其中 $\widetilde{\mathbf{E}}_n$ 表示对测度 $\widetilde{\mathbf{P}}_n$ 的数学期望.

对于现在考虑的情形, $\alpha_1 = 0$, $\alpha_n = \Delta g(n)$, $n \geqslant 2$, 因此对于 $n \geqslant 2$,

$$z_n^{-1} = \exp\left\{-\sum_{k=2}^{n} \Delta g(k) \cdot \xi_k + \frac{1}{2} \sum_{k=2}^{n} [\Delta g(k)]^2\right\}.$$

由分部求和公式 (见第四章 §3 引理 2 的证明), 得

$$\sum_{k=2}^{n}\Delta g(k)\cdot\xi_k=\Delta g(n)\cdot S_n-\frac{1}{2}\sum_{k=2}^{n}S_{k-1}\Delta[\Delta g(k)];$$

由此并注意到, 根据定理的条件 $\Delta g(k)\geqslant 0,\Delta[\Delta g(k)]\leqslant 0$, 可见在集合 $\{\tau>n\}=\{S_k\geqslant g(k),1\leqslant k\leqslant n\}$ 上, 有

$$\begin{aligned}\sum_{k=2}^{n}\Delta g(k)\cdot\xi_k&\geqslant\Delta g(n)\cdot g(n)-\sum_{k=3}^{n}g(k-1)\Delta[\Delta g(k)]-\xi_1\Delta g(2)\\&=\sum_{k=2}^{n}[\Delta g(k)]^2+g(1)\Delta g(2)-\xi_1\Delta g(2).\end{aligned}$$

因而, 由 (16) 式, 有

$$\begin{aligned}\mathbf{P}\{\tau>n\}&\leqslant\exp\left\{-\frac{1}{2}\sum_{k=2}^{n}[\Delta g(k)]^2-g(1)\Delta g(2)\right\}\widetilde{\mathbf{E}}_nI(\tau>n)\mathrm{e}^{-\xi_1\Delta g(2)}\\&=\exp\{-g(1)\Delta g(2)\}\exp\left\{-\frac{1}{2}\sum_{k=2}^{n}[\Delta g(k)]^2\right\}\widetilde{\mathbf{E}}_nI(\tau>n)\mathrm{e}^{-\xi_1\Delta g(2)},\end{aligned}$$

其中

$$\widetilde{\mathbf{E}}_nI(\tau>n)\mathrm{e}^{-\xi_1\Delta g(2)}\leqslant\mathbf{E}z_n\mathrm{e}^{-\xi_1\Delta g(2)}=\mathbf{E}\mathrm{e}^{-\xi_1\Delta g(2)}<\infty.$$

因此,

$$\mathbf{P}\{\tau>n\}\leqslant C\exp\left\{-\frac{1}{2}\sum_{k=2}^{n}[\Delta g(k)]^2\right\},$$

其中 C 是某一正常数, 于是证明了 (11) 式的上侧估计.

3. 双侧界限的情形 “测度绝对连续替换” 的思想, 亦可对于双侧边界的情形, 研究类似的问题. 我们 (不加证明地) 给出这一方向上的结果之一.

定理 2 设 $\xi_1,\xi_2,\cdots$ 是独立同分布的随机变量序列, $\xi_i\sim N(0,1)$. 假设 $f=f(n)$ 是正值函数, 满足条件:

$$f(n)\to\infty,\quad n\to\infty,$$

且

$$\sum_{k=2}^{n}[\Delta f(k)]^2=o\left(\sum_{k=1}^{n}f^{-2}(k)\right),\quad n\to\infty$$

那么, 如果

$$\sigma=\inf\{n\geqslant 1:|S_n|\geqslant f(n)\},$$

则

$$\mathbf{P}\{\sigma>n\}=\exp\left\{-\frac{\pi^2}{8}\sum_{k=1}^{n}f^{-2}(k)[1+o(1)]\right\},\quad n\to\infty.\tag{17}$$

4. 练习题

1. 证明由 (4) 式决定的序列是鞅. 问在不要求条件 $\mathbf{P}\{|\alpha_n| \leqslant c = 1, n \geqslant 1\}$ 成立的情况下, 这是否成立?

2. 证明公式 (13).

3. 证明公式 (17).

§8. 相依随机变量之和的中心极限定理

1. 函数的中心极限定理 在第三章 §4 中, 我们对于随机变量 $\xi_{1n}, \cdots, \xi_{nn}$ 之和 $S_n = \xi_{1n} + \cdots + \xi_{nn} (n \geqslant 1)$, 证明了中心极限定理, 当时假设: $\xi_{1n}, \cdots, \xi_{nn}$ 相互独立, 其二阶矩有限, 以及各被加项的极限可忽略性. 在这一节, 我们将不再假设满足独立性条件, 甚至也不再要求一阶绝对矩有限. 不过, 仍然假设被加项的极限可忽略性.

这样, 假设在概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 上给定随机序列

$$\xi^n = (\xi_{nk}, \mathscr{F}_k^n), \quad 0 \leqslant k \leqslant n, \quad n \geqslant 1,$$

其中 $\xi_{n0} = 0, \mathscr{F}_0^n = \{\varnothing, \Omega\}, \mathscr{F}_k^n \subseteq \mathscr{F}_{k+1}^n \subseteq \mathscr{F} (k+1 \leqslant n)$. 设

$$X_t^n = \sum_{k=0}^{[nt]} \xi_{nk}, \quad 0 \leqslant t \leqslant 1.$$

定理 1 设对于固定的 $0 < t \leqslant 1$, 满足下列条件: 对于任意 $\varepsilon \in (0,1]$, 当 $n \to \infty$ 时, 有

$$\text{(A)} \qquad \sum_{k=1}^{[nt]} \mathbf{P}\{|\xi_{nk}| > \varepsilon | \mathscr{F}_{k-1}^n\} \xrightarrow{\mathbf{P}} 0,$$

$$\text{(B)} \qquad \sum_{k=1}^{[nt]} \mathbf{E}[\xi_{nk} I(|\xi_{nk}| \leqslant \varepsilon | \mathscr{F}_{k-1}^n)] \xrightarrow{\mathbf{P}} 0,$$

$$\text{(C)} \qquad \sum_{k=1}^{[nt]} \mathbf{D}[\xi_{nk} I(|\xi_{nk}| \leqslant \varepsilon) | \mathscr{F}_{k-1}^n] \xrightarrow{\mathbf{P}} \sigma_t^2.$$

那么,

$$X_t^n \xrightarrow{d} N(0, \sigma_t^2).$$

注 1 条件 (A) 和 (B) 保障随机变量 X_t^n 可以表示为

$$X_t^n = Y_t^n + Z_t^n, \quad \text{而 } Z_t^n \xrightarrow{\mathbf{P}} 0, \quad Y_t^n = \sum_{k=0}^{[nt]} \eta_{nk},$$

其中序列 $\eta^n = (\eta_{nk}, \mathscr{F}^n_k)$ 是鞅 - 差 $\mathbf{E}(\eta_{nk}|\mathscr{F}^n_{k-1}) = 0$ 其中对 $0 \leqslant k \leqslant n, n \geqslant 1$ 一致 $|\eta_{nk}| \leqslant c$. 因此 (在所考虑的条件下) 定理的证明实际上归结为, 形成鞅 - 差序列的极限定理的证明.

对于随机变量 $\xi_{1n}, \cdots, \xi_{nn}$ 独立的情形, 条件 (A), (B) 和 (C) 当 $t = 1$ 时就是条件 $(\sigma^2 = \sigma_1^2)$:

$$\text{(a)} \qquad \sum_{k=1}^n \mathbf{P}\{|\xi_{nk}| > \varepsilon\} \to 0,$$

$$\text{(b)} \qquad \sum_{k=1}^n \mathbf{E}[\xi_{nk} I(|\xi_{nk}| \leqslant \varepsilon)] \to 0,$$

$$\text{(c)} \qquad \sum_{k=1}^n \mathbf{D}[\xi_{nk} I(|\xi_{nk}| \leqslant \varepsilon)] \to \sigma^2.$$

这由格涅坚科 (Б. В. Гнеденко) 和柯尔莫戈洛夫的专著 [16] 是熟知的. 因而由定理 1 得如下推论.

系 如果 $\xi_{1n}, \cdots, \xi_{nn} (n \geqslant 1)$ 是独立随机变量, 则

$$\text{(a)}, \text{(b)}, \text{(c)} \Rightarrow X_1^n = \sum_{k=1}^n \xi_{nk} \xrightarrow{d} N(0, \sigma^2).$$

注 2 在条件 (C) 中不排除 $\sigma_t^2 = 0$ 的情形. 这样, 定理 1 中也包括收敛于退化分布的情形 $(X_t^n \xrightarrow{d} 0)$.

注 3 用定理 1 的方法可以表述和证明更加一般的命题.

假设 $0 < t_1 < t_2 < \cdots < t_j \leqslant 1, \sigma_{t_1} \leqslant \sigma_{t_2} \leqslant \cdots \leqslant \sigma_{t_j}$, 而 $\varepsilon_1, \cdots, \varepsilon_j$ 是均值为 0 的独立高斯随机变量, 且 $\mathbf{E}\varepsilon_k^2 = \sigma_{t_k}^2 - \sigma_{t_{k-1}}^2$. 建立高斯随机向量

$$(W_{t_1}, \cdots, W_{t_j}), \quad W_{t_k} = \varepsilon_1 + \cdots + \varepsilon_k.$$

假如条件 (A),(B) 和 (C) 当 $t = t_1, \cdots, t_j$ 时成立, 则随机变量 $X_{t_1}^n, \cdots, X_{t_j}^n$ 的联合分布 $P_{t_1,\cdots,t_j}^n$ 弱收敛于随机变量 $(W_{t_1}, \cdots, W_{t_j})$ 的高斯分布 $P_{t_1,\cdots,t_j}$: $P_{t_1,\cdots,t_j}^n \xrightarrow{w} P_{t_1,\cdots,t_j}$.

注 4 设 $(\sigma_t^2)_{0 \leqslant t \leqslant 1}$ 是连续不减函数, $\sigma_t^2 = 0$. 以 $W = (W_t)_{0 \leqslant t \leqslant 1}$ 表示布朗运动过程 (维纳过程), 且 $\mathbf{E}W_t = 0, \mathbf{E}W_t^2 = \sigma_t^2$. 在第二章 §13 中对于 $\sigma_t^2 = t$ 定义了该过程. 在没有这一条件的情况下, 该过程可以类似地定义为, 具有独立增量的高斯过程 $W = (W_t)_{0 \leqslant t \leqslant 1}$, 且 $W_0 = 0$, 而其协方差函数为 $r(s,t) = \min(\sigma_s^2, \sigma_t^2)$. 在随机过程的一般理论中, 证明具有连续轨道的这样的过程总是存在的. (当 $\sigma_t^2 = t$ 时这样的过程称为标准布朗运动).

如果以 P^n 和 P 表示过程 X^n 和 W 在函数空间 $(D, \mathscr{B}(D))$ (见第二章 §2 第 7 小节) 中的概率分布, 则可以断定: 对于一切 $0 < t \leqslant 1$ 成立的条件 (A), (B) 和 (C),

不仅可以保证上述有限维分布的弱收敛 ($P^n_{t_1,\cdots,t_j} \xrightarrow{w} P_{t_1,\cdots,t_j}, t_1 < t_2 < \cdots < t_j \leqslant t, j = 1,2,\cdots$), 而且也保证函数的收敛性, 即过程 X^n 的分布 P^n 弱收敛于过程 W 的分布 P. (有关细节, 参见 [5], [91], [87]). 这一结果, 一般称做函数的中心极限定理, 或者 (对于 $n \geqslant 1$, 当随机变量 $\xi_{1n},\cdots,\xi_{nn}$ 独立时) 称做 (汤斯凯 [M. Donsker] – 普罗霍罗夫) 不变原理.

2. 一致渐近可忽略性

定理 2 (1) 条件 (A) 等价于一致极限可忽略条件 (一致渐近可忽略性)

(A*) $$\max_{1\leqslant k\leqslant [nt]} |\xi_{nk}| \xrightarrow{\mathbf{P}} 0.$$

(2) 在条件 (A) 或 (A*) 下, 条件 (C) 等价于条件

(C*) $$\sum_{k=1}^{[nt]} [\xi_{nk} - \mathbf{E}\xi_{nk} I(|\xi_{nk}| \leqslant 1)|\mathscr{F}^n_{k-1}]^2 \xrightarrow{\mathbf{P}} \sigma^2_t.$$

((A*) 和 (C*) 中与 (A) 和 (C) 中 t 的值相同.)

定理 3 假设对于每一个 $n \geqslant 1$, 序列

$$\xi^n = (\xi_{nk}, \mathscr{F}^n_k), \quad 0 \leqslant k \leqslant n,$$

是平方可积鞅 — 差, 即 $\mathbf{E}\xi^2_{nk} < \infty, \mathbf{E}(\xi_{nk}|\mathscr{F}^n_{k-1}) = 0$.

假如满足林德伯格 (J. W. Lindeberg) 条件: 对于任意 $\varepsilon > 0$, 有

(L) $$\sum_{k=1}^{[nt]} \mathbf{E}[\xi^2_{nk} I(|\xi_{nk}| \geqslant \varepsilon)|\mathscr{F}^n_{k-1}] \xrightarrow{\mathbf{P}} 0.$$

那么, 条件 (C) 等价于条件

$$\langle X^n \rangle_t \xrightarrow{\mathbf{P}} \sigma^2_t, \tag{1}$$

其中

$$\langle X^n \rangle_t = \sum_{k=1}^{[nt]} \mathbf{E}(\xi^2_{nk}|\mathscr{F}^n_{k-1}) \tag{2}$$

是二次特征, 而条件 (C*) 等价于条件

$$[X^n]_t \xrightarrow{\mathbf{P}} \sigma^2_t, \tag{3}$$

其中

$$[X^n]_t = \sum_{k=1}^{[nt]} \xi^2_{nk} \tag{4}$$

是二次特征.

由定理 1~3, 得下面的定理.

定理 4 假设对于平方可积鞅 - 差 $\xi^n=(\xi_{nk},\mathscr{F}^n_k), n\geqslant 1$ 和 (对给定的 $0<t\leqslant 1$), 满足条件林德伯格 (L), 那么,

$$\sum_{k=1}^{[nt]}\mathbf{E}(\xi_{nk}^2|\mathscr{F}^n_{k-1})\xrightarrow{\mathbf{P}}\sigma_t^2\Rightarrow X_t^n\xrightarrow{d}N(0,\sigma_t^2),\tag{5}$$

$$\sum_{k=1}^{[nt]}\xi_{nk}^2\xrightarrow{\mathbf{P}}\sigma_t^2\Rightarrow X_t^n\xrightarrow{d}N(0,\sigma_t^2).\tag{6}$$

3. 定理 1 的证明 1) 将 X_t^n 表示为如下形式:

$$\begin{aligned}X_t^n&=\sum_{k=1}^{[nt]}\xi_{nk}I(|\xi_{nk}|\leqslant 1)+\sum_{k=1}^{[nt]}\xi_{nk}I(|\xi_{nk}|>1)\\&=\sum_{k=1}^{[nt]}\mathbf{E}[\xi_{nk}I(|\xi_{nk}|\leqslant 1)|\mathscr{F}^n_{k-1}]+\sum_{k=1}^{[nt]}\xi_{nk}I(|\xi_{nk}|>1)\\&\quad+\sum_{k=1}^{[nt]}\{\xi_{nk}I(|\xi_{nk}|\leqslant 1)-\mathbf{E}[\xi_{nk}I(|\xi_{nk}|\leqslant 1)|\mathscr{F}^n_{k-1}]\}.\end{aligned}\tag{7}$$

记

$$\begin{gathered}B_t^n=\sum_{k=1}^{[nt]}\mathbf{E}[\xi_{nk}I(|\xi_{nk}|\leqslant 1)|\mathscr{F}^n_{k-1}],\\ \mu_k^n(\Gamma)=I(\xi_{nk}\in\Gamma),\\ \nu_k^n(\Gamma)=\mathbf{P}(\xi_{nk}\in\Gamma|\mathscr{F}^n_{k-1}),\end{gathered}\tag{8}$$

其中 Γ 是属于集系 $\mathscr{A}_0$ 生成的最小 $\sigma-$ 代数 $\mathscr{B}_0=\sigma(\mathscr{A}_0)$ 的集合, 而 $\mathscr{A}_0$ 是 $\mathbb{R}_0=\mathbb{R}\backslash\{0\}$ 中有限个形如 $(a,b]$ (不包含点 $\{0\}$) 的、不相交区间之和形成的集系; $\mathbf{P}(\xi_{nk}\in\Gamma|\mathscr{F}^n_{k-1})$ 是 ξ_{nk} 关于 $\sigma-$ 代数 $\mathscr{F}^n_{k-1}$ 的正则条件分布.

那么, (7) 式可以写成

$$X_t^n=B_t^n+\sum_{k=1}^{[nt]}\int_{\{|x|>1\}}x\mathrm{d}\mu_k^n+\sum_{k=1}^{[nt]}\int_{\{|x|\leqslant 1\}}x\mathrm{d}(\mu_k^n-\nu_k^n).\tag{9}$$

表达式 (9) 称做序列 $(X_t^n,\mathscr{F}^n_{[nt]})$ 的*典范分解式*. (所有积分应理解为对于每一个基本事件由定义, 勒贝格 — 斯蒂尔切斯积分.)

根据条件 (B), $B_t^n\xrightarrow{\mathbf{P}}0$. 现在证明, 由于条件 (A), 有

$$\sum_{k=1}^{[nt]}\int_{\{|x|>1\}}|x|\mathrm{d}\mu_k^n\xrightarrow{\mathbf{P}}0.\tag{10}$$

有

$$\sum_{k=1}^{[nt]} \int_{\{|x|>1\}} |x| \mathrm{d}\mu_k^n = \sum_{k=1}^{[nt]} |\xi_{nk}| I(|\xi_{nk}| > 1). \tag{11}$$

对于任意 $\delta \in (0,1)$,

$$\left\{ \sum_{k=1}^{[nt]} |\xi_{nk}| I(|\xi_{nk}| > 1) > \delta \right\} \subseteq \left\{ \sum_{k=1}^{[nt]} I(|\xi_{nk}| > 1) > \delta \right\}. \tag{12}$$

显然,

$$\sum_{k=1}^{[nt]} I(|\xi_{nk}| > 1) = \sum_{k=1}^{[nt]} \int_{\{|x|>1\}} \mathrm{d}\mu_k^n \ (\equiv U_{[nt]}^n).$$

由条件 (A), 有

$$V_{[nt]}^n \equiv \sum_{k=1}^{[nt]} \int_{\{|x|>1\}} \mathrm{d}\nu_k^n \xrightarrow{\mathbf{P}} 0, \tag{13}$$

并且 $V_{[nt]}^n$ 是 $\mathscr{F}_{k-1}^n$- 可测的.

那么, 由 §3 定理 4 的系, 有

$$V_{[nt]}^n \xrightarrow{\mathbf{P}} 0 \Rightarrow U_{[nt]}^n \xrightarrow{\mathbf{P}} 0. \tag{14}$$

注意, 由 §3 定理 4 的系, 以及不等式 $\Delta U_{[nt]}^n \leqslant 1$, 有相反的蕴含关系

$$U_{[nt]}^n \xrightarrow{\mathbf{P}} 0 \Rightarrow V_{[nt]}^n \xrightarrow{\mathbf{P}} 0, \tag{15}$$

这在证明定理 2 时将要用到.

由 (11)~(14) 式的所要证明的 (10) 式.

这样,

$$X_t^n = Y_t^n + Z_t^n, \tag{16}$$

其中

$$Y_t^n = \sum_{k=1}^{[nt]} \int_{\{|x|\leqslant 1\}} x \mathrm{d}(\mu_k^n - \nu_k^n), \tag{17}$$

而

$$Z_t^n = B_t^n + \sum_{k=1}^{[nt]} \int_{\{|x|>1\}} x \mathrm{d}\mu_k^n \xrightarrow{\mathbf{P}} 0. \tag{18}$$

2) 由于练习题 1, 由此可见为证明收敛性 $X_t^n \xrightarrow{d} N(0, \sigma_t^2)$, 只需证明

$$Y_t^n \xrightarrow{d} N(0, \sigma_t^2). \tag{19}$$

把 Y_t^n 表示为 $(\varepsilon \in (0,1])$

$$Y_t^n = \gamma_{[nt]}^n(\varepsilon) + \Delta_{[nt]}^n(\varepsilon),$$

其中

$$\gamma_{[nt]}^n(\varepsilon) = \sum_{k=1}^{[nt]} \int_{\{\varepsilon<|x|\leqslant 1\}} x\mathrm{d}(\mu_k^n - \nu_k^n), \tag{20}$$

$$\Delta_{[nt]}^n(\varepsilon) = \sum_{k=1}^{[nt]} \int_{\{|x|\leqslant \varepsilon\}} x\mathrm{d}(\mu_k^n - \nu_k^n), \tag{21}$$

像 (10) 式的证明一样, 由于条件 (A) 容易证明 $\gamma_{[nt]}^n(\varepsilon) \xrightarrow{\mathbf{P}} 0, n\to\infty$. 序列 $\Delta^n(\varepsilon) = (\Delta_k^n(\varepsilon), \mathscr{F}_k^n), 1\leqslant k\leqslant n$ 是平方可积鞅, 且具有二次特征

$$\begin{aligned}\langle \Delta^n(\varepsilon)\rangle_k &= \sum_{i=1}^k \left[\int_{\{|x|\leqslant\varepsilon\}} x^2\mathrm{d}\nu_i^n - \left(\int_{\{|x|\leqslant\varepsilon\}} x\mathrm{d}\nu_i^n\right)^2\right] \\ &= \sum_{i=1}^k \mathbf{D}[\xi_{ni} I(|\xi_{ni}|\leqslant \varepsilon|\mathscr{F}_{i-1}^n)].\end{aligned}$$

由于条件 (C), 有

$$\langle \Delta^n(\varepsilon)\rangle_{[nt]} \xrightarrow{\mathbf{P}} \sigma_t^2.$$

从而, 对于任意 $\varepsilon\in(0,1]$

$$\max\{\gamma_{[nt]}^n(\varepsilon), |\langle\Delta^n(\varepsilon)\rangle_{[nt]} - \sigma_t^2|\} \xrightarrow{\mathbf{P}} 0.$$

根据练习题 2, 存在数列 $\varepsilon_n \downarrow 0$, 使

$$\gamma_{[nt]}^n(\varepsilon_n) \xrightarrow{\mathbf{P}} 0, \quad \langle \Delta^n(\varepsilon_n)\rangle_{[nt]} \xrightarrow{\mathbf{P}} \sigma_t^2.$$

因此, 仍然由于练习题 1, 只需证明,

$$M_{[nt]}^n \xrightarrow{d} N(0,\sigma_t^2). \tag{22}$$

其中

$$M_k^n = \Delta_k^n(\varepsilon_n) = \sum_{i=1}^k \int_{\{|x|\leqslant\varepsilon_n\}} x\mathrm{d}(\mu_i^n - \nu_i^n). \tag{23}$$

假设对于 $\Gamma\in\mathscr{B}_0$,

$$\widetilde{\mu}_k^n(\Gamma) = I(\Delta M_k^n \in \Gamma), \quad \widetilde{\nu}_k^n(\Gamma) = \mathbf{P}(\Delta M_k^n\in\Gamma|\mathscr{F}_{k-1}^n)$$

是正则条件概率, $\Delta M_k^n = M_k^n - M_{k-1}^n, k \geqslant 1, M_0^n = 0$. 那么, 平方可积鞅 $M^n = (M_k^n, \mathscr{F}_k^n)$, $1 \leqslant k \leqslant n$, 显然可以写成下面的形式:

$$M_k^n = \sum_{i=1}^k \Delta M_i^n = \sum_{i=1}^k \int_{\{|x| \leqslant 2\varepsilon_n\}} x \mathrm{d}\widetilde{\mu}_i^n.$$

(注意, 由于 (23), 可见 $|\Delta M_i^n| \leqslant 2\varepsilon_n$.)

3) 根据第三章 §3 定理 1, 为证明 (22) 式, 需要对于任意实数 λ, 证明

$$\mathbf{E} \mathrm{e}^{i\lambda M_{[nt]}^n} \to \mathrm{e}^{-\frac{\lambda^2 \sigma_t^2}{2}}. \tag{24}$$

记

$$G_k^n = \sum_{j=1}^k \int_{\{|x| \leqslant 2\varepsilon_n\}} (\mathrm{e}^{i\lambda x} - 1) \mathrm{d}\widetilde{\nu}_j^n$$

和

$$\mathscr{E}_t^n(G^n) = \prod_{j=1}^k (1 + \Delta G_j^n).$$

注意到

$$1 + \Delta G_k^n = 1 + \int_{\{|x| \leqslant 2\varepsilon_n\}} (\mathrm{e}^{i\lambda x} - 1) \mathrm{d}\widetilde{\nu}_k^n = \mathbf{E}(\mathrm{e}^{i\lambda \Delta M_k^n} | \mathscr{F}_{k-1}^n),$$

从而

$$\mathscr{E}_k^n(G^n) = \prod_{j=1}^k \mathbf{E}(\mathrm{e}^{i\lambda \Delta M_j^n} | \mathscr{F}_{j-1}^n).$$

4) 根据第 4 小节已证明的引理 4, 为验证 (24) 式只需证明, 对于任意实数 λ, 有

$$|\mathscr{E}_{[nt]}^n(G^n)| = \left| \prod_{j=1}^{[nt]} \mathbf{E}(\mathrm{e}^{i\lambda \Delta M_j^n} | \mathscr{F}_{j-1}^n) \right| \geqslant c(\lambda) > 0 \tag{25}$$

和

$$|\mathscr{E}_{[nt]}^n(G^n)| \xrightarrow{\mathbf{P}} \mathrm{e}^{-\frac{\lambda^2 \sigma_t^2}{2}}. \tag{26}$$

为此, 将 $\mathscr{E}_k^n(G^n)$ 表示为

$$\mathscr{E}_k^n(G^n) = \mathrm{e}^{G_k^n} \prod_{j=1}^k (1 + \Delta G_j^n) \mathrm{e}^{-\Delta G_j^n}.$$

(对照第二章 §6 中由 (76) 式定义的函数 $\mathscr{E}_t(A)$).

由于

$$\int_{\{|x| \leqslant 2\varepsilon_n\}} x \mathrm{d}\widetilde{\nu}_j^n = \mathbf{E}(\Delta M_j^n | \mathscr{F}_{j-1}^n) = 0,$$

可见

$$G_k^n = \sum_{j=1}^{k} \int_{\{|x| \leqslant 2\varepsilon_n\}} (\mathrm{e}^{i\lambda x} - 1 - i\lambda x) \mathrm{d}\widetilde{\nu}_j^n. \tag{27}$$

从而,

$$|\Delta G_k^n| \leqslant \int_{\{|x| \leqslant 2\varepsilon_n\}} |\mathrm{e}^{i\lambda x} - 1 - i\lambda x| \mathrm{d}\widetilde{\nu}_k^n \leqslant \frac{\lambda^2}{2} \int_{\{|x| \leqslant 2\varepsilon_n\}} x^2 \mathrm{d}\widetilde{\nu}_k^n \leqslant \frac{\lambda^2}{2}(2\varepsilon_n)^2 \to 0 \tag{28}$$

和

$$\sum_{j=1}^{k} |\Delta G_j^n| \leqslant \frac{\lambda^2}{2} \sum_{j=1}^{k} \int_{\{|x| \leqslant 2\varepsilon_n\}} x^2 \mathrm{d}\widetilde{\nu}_j^n = \frac{\lambda^2}{2} \langle M^n \rangle_k. \tag{29}$$

根据条件 (C),

$$\langle M^n \rangle_{[nt]} \xrightarrow{\mathbf{P}} \sigma_t^2. \tag{30}$$

首先, 设 $\langle M^n \rangle_{[nt]} \leqslant a$ ($\mathbf{P}$-a.c.), 则由 (28), (29) 式以及练习题 3, 显然

$$\prod_{k=1}^{[nt]} (1 + \Delta G_k^n) \mathrm{e}^{-\Delta G_k^n} \xrightarrow{\mathbf{P}} 1, \quad n \to \infty,$$

因此, 为证明 (26) 式, 只需证明

$$G_{[nt]}^n \xrightarrow{\mathbf{P}} -\frac{\lambda^2 \sigma_t^2}{2} \tag{31}$$

或 (由 (27), (29) 和 (30) 各式)

$$\sum_{k=1}^{[nt]} \int_{\{|x| \leqslant 2\varepsilon_n\}} \left(\mathrm{e}^{i\lambda x} - 1 - i\lambda x + \frac{\lambda^2 x^2}{2} \right) \mathrm{d}\widetilde{\nu}_k^n \xrightarrow{\mathbf{P}} 0. \tag{32}$$

由于

$$\left| \mathrm{e}^{i\lambda x} - 1 - i\lambda x + \frac{\lambda^2 x^2}{2} \right| \leqslant \frac{|\lambda x|^3}{6},$$

可见

$$\begin{aligned} \sum_{k=1}^{[nt]} \int_{\{|x| \leqslant 2\varepsilon_n\}} \left| \mathrm{e}^{i\lambda x} - 1 - i\lambda x + \frac{\lambda^2 x^2}{2} \right| \mathrm{d}\widetilde{\nu}_k^n &\leqslant \frac{|\lambda|^3}{6} \times 2\varepsilon_n \times \sum_{k=1}^{[nt]} \int_{\{|x| \leqslant 2\varepsilon_n\}} x^2 \mathrm{d}\widetilde{\nu}_k^n \\ &= \frac{|\lambda|^3 \varepsilon_n}{3} \langle M^n \rangle_{[nt]} \leqslant \frac{|\lambda|^3 \varepsilon_n}{3} a \to 0, \quad n \to \infty. \end{aligned}$$

这样, 假如 $\langle M^n \rangle_{[nk]} \leqslant a$ ($\mathbf{P}$-a.c.), 则 (31) 式成立, 从而 (26) 式得证.

5) 现在证明性质 (25). 因为

$$|\mathrm{e}^{i\lambda x} - 1 - i\lambda x| \leqslant \frac{(\lambda x)^2}{2},$$

所以, 由 (28) 式可见, 对于充分大的 n, 有

$$
\begin{aligned}
|\mathscr{E}_k^n(G^n)| = \left|\prod_{j=1}^{k}(1+\Delta G_j^n)\right| &\geqslant \prod_{j=1}^{k}\left(1-\frac{\lambda^2}{2}\Delta\langle M^n\rangle_j\right)\\
&= \exp\left\{\sum_{j=1}^{k}\ln\left(1-\frac{\lambda^2}{2}\Delta\langle M^n\rangle_j\right)\right\}.
\end{aligned}
$$

而 (对于充分大的 n)

$$
\ln\left(1-\frac{\lambda^2}{2}\Delta\langle M^n\rangle_j\right) \geqslant -\frac{\frac{\lambda^2}{2}\Delta\langle M^n\rangle_j}{1-\frac{\lambda^2}{2}\Delta\langle M^n\rangle_j},
$$

其中 $\Delta\langle M^n\rangle_j \leqslant (2\varepsilon_n)^2 \downarrow 0, n\to\infty$, 因此存在 $n_0 = n_0(\lambda)$, 使对于一切 $n \geqslant n_0(\lambda)$, 有

$$
|\mathscr{E}_k^n(G^n)| \geqslant \mathrm{e}^{-\lambda^2\langle M^n\rangle_k},
$$

从而

$$
|\mathscr{E}_{[nt]}^n(G^n)| \geqslant \mathrm{e}^{-\lambda^2\langle M^n\rangle_{[nt]}} \geqslant \mathrm{e}^{-\lambda^2 a}.
$$

于是, 在 $\langle M^n\rangle_{[nk]} \leqslant a$ ($\mathbf{P}$–a.c.) 的条件下, 定理得证,

6) 为去掉条件 $\langle M^n\rangle_{[nk]} \leqslant a$ ($\mathbf{P}$–a.c.), 我们作如下处理, 设

$$
\tau^n = \begin{cases} \inf\{k\leqslant [nt] : \langle M^n\rangle_k \geqslant \sigma_t^2+1\}, & 若\ \langle M^n\rangle_{[nt]} \geqslant \sigma_t^2+1,\\ \infty, & 若\ \langle M^n\rangle_{[nt]} < \sigma_t^2+1. \end{cases}
$$

那么, 对于 $\overline{M}^n = M_{k\wedge\tau^n}^n$, 有

$$
\langle\overline{M}^n\rangle_{[nt]} = M_{[nt]\wedge\tau^n} \leqslant 1+\sigma_t^2+2\varepsilon_n^2 \leqslant 1+\sigma_t^2+2\varepsilon_1^2 (=a),
$$

且根据已证明的 (见 (24) 式)

$$
\mathbf{E}\mathrm{e}^{i\lambda\overline{M}_{[nt]}^n} \to \mathrm{e}^{-\frac{\lambda^2\sigma_t^2}{2}}.
$$

但是

$$
\lim_n |\mathbf{E}(\mathrm{e}^{i\lambda M_{[nt]}^n} - \mathrm{e}^{i\lambda\overline{M}_{[nt]}^n})| \leqslant 2\lim_n \mathbf{P}\{\tau^n<\infty\} = 0.
$$

因此,

$$
\lim_n \mathbf{E}\mathrm{e}^{i\lambda M_{[nt]}^n} = \lim_n \mathbf{E}(\mathrm{e}^{i\lambda M_{[nt]}^n} - \mathrm{e}^{i\lambda\overline{M}_{[nt]}^n}) + \lim_n \mathbf{E}\mathrm{e}^{i\lambda\overline{M}_{[nt]}^n} = \mathrm{e}^{-\frac{\lambda^2\sigma_t^2}{2}}. \qquad \square
$$

注 为证明定理 1 的注 2 中提出的论断, 需要 (根据克拉默 - 沃尔德方法 [5]) 证明, 对于任意实数 $\lambda_1, \cdots, \lambda_j$, 证明

$$\mathbf{E}\exp\left\{i\left[\lambda_1 M^n_{[nt_1]} + \sum_{k=2}^{j}\lambda_k(M^n_{[nt_k]} - M^n_{[nt_{k-1}]})\right]\right\}$$
$$\to \exp\left\{-\frac{\lambda_1^2\sigma_{t_1}^2}{2} - \sum_{k=2}^{j}\frac{\lambda_k^2(\sigma_{t_k}^2 - \sigma_{t_{k-1}}^2)}{2}\right\}.$$

该式的证明与 (24) 式一样, 只是用平方可积鞅 $(\widehat{M}^n_k, \mathscr{F}^n_k)$ 代替 $(M^n_k, \mathscr{F}^n_k)$, 其中

$$\widehat{M}^n_k = \sum_{i}^{k}\nu_i \Delta M^n_i,$$

而

$$\nu_i = \begin{cases}\lambda_1, & 若\ i \leqslant [nt_1], \\ \lambda_k, & 若\ [nt_{k-1}] < i \leqslant [nt_k], 2 \leqslant k \leqslant j.\end{cases}$$

4. 辅助命题 在这一小节将证明一个简单的引理, 它可以将验证 (25) 式和 (26) 式, 归结为验证 (24) 式.

设 $\eta^n = (\eta_{nk}, \mathscr{F}^n_k), 1 \leqslant k \leqslant n, n \geqslant 1$ 是随机序列, $Y^n = \sum\limits_{k=1}^{n}\eta_{nk}$,

$$\mathscr{E}^n(\lambda) = \prod_{k=1}^{n}\mathbf{E}(e^{i\lambda\eta_{nk}} | \mathscr{F}^n_{k-1}), \quad \lambda \in \mathbb{R},$$

Y 是随机变量, 其中

$$\mathscr{E}^n(\lambda) = \mathbf{E}e^{i\lambda Y}, \quad \lambda \in \mathbb{R}.$$

引理 *如果 (对于给定的 λ) $|\mathscr{E}^n(\lambda)| \geqslant c(\lambda) > 0, n \geqslant 1$, 则收敛*

$$\mathbf{E}e^{i\lambda Y^n} \to \mathbf{E}e^{i\lambda Y} \tag{33}$$

的充分条件为收敛

$$\mathscr{E}^n(\lambda) \xrightarrow{\mathbf{P}} \mathscr{E}(\lambda). \tag{34}$$

证明 设

$$m^n(\lambda) = \frac{e^{i\lambda Y^n}}{\mathscr{E}^n(\lambda)},$$

则 $|m^n(\lambda)| \leqslant c^{-1}(\lambda) < \infty$, 且容易验证

$$\mathbf{E}m^n(\lambda) = 1.$$

因此, 由 (34) 式以及勒贝格控制收敛定理, 有

$$|\mathbf{E}e^{i\lambda Y^n} - \mathbf{E}e^{i\lambda Y}| = |\mathbf{E}(e^{i\lambda Y^n} - \mathscr{E}(\lambda))| \leqslant |\mathbf{E}\{m^n(\lambda)[\mathscr{E}^n(\lambda) - \mathscr{E}(\lambda)]\}|$$
$$\leqslant c^{-1}(\lambda)\mathbf{E}|\mathscr{E}^n(\lambda) - \mathscr{E}(\lambda)| \to 0, \quad n \to \infty. \qquad \square$$

注　由 (33) 式和假设 $|\mathscr{E}^n(\lambda)| \geqslant c(\lambda) > 0$, 可见 $\mathscr{E}(\lambda) \neq 0$. 实际上, 在不要求满足条件 $|\mathscr{E}^n(\lambda)| \geqslant c(\lambda) > 0$ 的情况下, 引理的论点仍然成立; 具体地可以表述为: 如果 $\mathscr{E}^n(\lambda) \xrightarrow{\mathbf{P}} \mathscr{E}(\lambda)$ 且 $\mathscr{E}(\lambda) \neq 0$, 则收敛性 (33) 式成立 (练习题 5).

5. 定理 2 的证明　1) 设 $\varepsilon > 0, \delta \in (0, \varepsilon)$, 并且为简便计设 $t = 1$. 由于

$$\max_{1 \leqslant k \leqslant n} |\xi_{nk}| \leqslant \varepsilon + \sum_{k=1}^{n} |\xi_{nk}| I(|\xi_{nk}| > \varepsilon).$$

和

$$\left\{\sum_{k=1}^{n} |\xi_{nk}| I(|\xi_{nk}| > \varepsilon) > \delta\right\} \subseteq \left\{\sum_{k=1}^{n} I(|\xi_{nk}| > \varepsilon) > \delta\right\},$$

则

$$\mathbf{P}\left\{\max_{1 \leqslant k \leqslant n} |\xi_{nk}| > \varepsilon + \delta\right\} \leqslant \mathbf{P}\left\{\sum_{k=1}^{n} I(|\xi_{nk}| > \varepsilon) > \delta\right\} = \mathbf{P}\left\{\sum_{k=1}^{n} \int_{\{|x| > \varepsilon\}} \mathrm{d}\mu_k^n > \delta\right\}.$$

如果满足条件 (A), 即当 $n \to \infty$ 时

$$\mathbf{P}\left\{\sum_{k=1}^{n} \int_{\{|x| > \varepsilon\}} \mathrm{d}\nu_k^n > \delta\right\} \to 0$$

则 (对照 (10) 式)

$$\mathbf{P}\left\{\sum_{k=1}^{n} \int_{\{|x| > \varepsilon\}} \mathrm{d}\mu_k^n > \delta\right\} \to 0.$$

从而 (A) $\Rightarrow$ (A*).

相反, 假设条件 (A*) 成立. 对于任意 $\varepsilon > 0$, 设

$$\sigma_n = \begin{cases} \min\left\{k \leqslant n : |\xi_{nk}| > \dfrac{\varepsilon}{2}\right\}, & \text{若} \ \max\limits_{1 \leqslant k \leqslant n} |\xi_{nk}| \geqslant \dfrac{\varepsilon}{2}, \\ \infty, & \text{若} \ \max\limits_{1 \leqslant k \leqslant n} |\xi_{nk}| < \dfrac{\varepsilon}{2}. \end{cases}$$

由 (A*) 式 $\lim\limits_{n} \mathbf{P}\{\sigma_n < \infty\} = 0$.

我们现在指出, 对于任意 $\delta \in (0, 1)$, 集合

$$\left\{\sum_{k=1}^{n \wedge \sigma_n} I\left(|\xi_{nk}| \geqslant \frac{\varepsilon}{2}\right) > \delta\right\} \quad \text{和} \quad \left\{\max_{1 \leqslant k \leqslant n \wedge \sigma_n} |\xi_{nk}| \geqslant \frac{\varepsilon}{2}\right\}$$

相等, 而根据条件 (A*), 有

$$\sum_{k=1}^{n \wedge \sigma_n} I\left(|\xi_{nk}| \geqslant \frac{\varepsilon}{2}\right) = \sum_{k=1}^{n \wedge \sigma_n} \int_{\left\{|x| \geqslant \frac{\varepsilon}{2}\right\}} \mathrm{d}\mu_k^n \xrightarrow{\mathbf{P}} 0.$$

因此, 由于 (15) 式, 有

$$\sum_{k=1}^{n\wedge\sigma_n}\int_{\{|x|\geqslant\varepsilon\}}\mathrm{d}\nu_k^n\leqslant\sum_{k=1}^{n\wedge\sigma_n}\int_{\{|x|\geqslant\frac{\varepsilon}{2}\}}\mathrm{d}\nu_k^n\xrightarrow{\mathbf{P}}0;$$

由该式连同性质 $\lim\limits_n\mathbf{P}\{\sigma_n<\infty\}=0$ 最后证明蕴涵关系 $(\mathrm{A}^*)\Rightarrow(\mathrm{A})$.

2) 仍然假设 $t=1$. 固定某个 $\varepsilon\in(0,1]$, 对于任意 $\delta\in(0,\varepsilon]$, 考虑平方可积鞅 (见 (21) 式),

$$\Delta^n(\delta)=(\Delta_k^n(\delta),\mathscr{F}_k^n),\quad 1\leqslant k\leqslant n.$$

由于条件 (C), 对于给定的 $\varepsilon\in(0,1]$, 有

$$\langle\Delta^n(\varepsilon)\rangle_n\xrightarrow{\mathbf{P}}\sigma_1^2.$$

从而, 由于条件 (A), 容易导出, 对于任意 $\delta\in(0,\varepsilon]$, 有

$$\langle\Delta^n(\delta)\rangle_n\xrightarrow{\mathbf{P}}\sigma_1^2. \tag{35}$$

现在证明, 由条件 (C^*) 和 (A), 或等价地, 由条件 (C^*) 和 (A^*) 可见, 对于任意 $\delta\in(0,\varepsilon]$, 有

$$[\Delta^n(\delta)]_n\xrightarrow{\mathbf{P}}\sigma_1^2, \tag{36}$$

其中

$$[\Delta^n(\delta)]_n=\sum_{k=1}^{n}\left[\xi_{nk}I(|\xi_{nk}|\leqslant\delta)-\int_{\{|x|\leqslant\delta\}}x\mathrm{d}\nu_k^n\right]^2.$$

事实上, 由于 (A) 式, 容易验证,

$$[\Delta^n(\delta)]_n-[\Delta^n(1)]_n\xrightarrow{\mathbf{P}}0. \tag{37}$$

$$\begin{aligned}&\left|\sum_{k=1}^{n}\left[\xi_{nk}-\int_{\{|x|\leqslant1\}}x\mathrm{d}\nu_k^n\right]^2-\sum_{k=1}^{n}\left[\xi_{nk}I(|\xi_{nk}|\leqslant1)-\int_{\{|x|\leqslant1\}}x\mathrm{d}\nu_k^n\right]^2\right|\\&\leqslant\sum_{k=1}^{n}I(|\xi_{nk}|>1)\left[\xi_{nk}^2+2|\xi_{nk}|\times\left|\int_{\{|x|\leqslant1\}}x\mathrm{d}(\mu_k^n-\nu_k^n)\right|\right]\\&\leqslant5\sum_{k=1}^{n}I(|\xi_{nk}|>1)\xi_{nk}^2\leqslant5\max_{1\leqslant k\leqslant n}\xi_{nk}^2\times\sum_{k=1}^{n}\int_{\{|x|>1\}}\mathrm{d}\mu_k^n\xrightarrow{\mathbf{P}}0.\end{aligned} \tag{38}$$

因此, 由 (37) 式和 (38) 式得 (36) 式.

这样, 为证明条件 (C) 与条件 (C^*) 等价, 只需证明: 在 (对于给定的 $\varepsilon\in(0,1]$) 条件 (C) 满足时, 同时, 若对于任意 $a>0$, 条件 (C^*) 成立, 则

$$\lim_{\delta\to0}\varlimsup_n\mathbf{P}\{|[\Delta^n(\delta)]_n-\langle\Delta^n(\delta)\rangle_n|>a\}=0. \tag{39}$$

记 $m_k^n(\delta) = [\Delta^n(\delta)]_k - \langle\Delta^n(\delta)\rangle_k, 1 \leqslant k \leqslant n$. 序列 $m^n(\delta) = (m_k^n(\delta), \mathscr{F}_k^n)$ 是平方可积鞅, 这时序列 $(m^n(\delta))^2$ 被序列 $[\Delta^n(\delta)]_n$ 和 $\langle\Delta^n(\delta)\rangle_n$ 控制 (由 §3 定义).

显然,

$$\begin{aligned}[m^n(\delta)]_n = \sum_{k=1}^{n}[\Delta m_k^n(\delta)]^2 &\leqslant \max_{1\leqslant k\leqslant n}|\Delta m_k^n(\delta)| \times \{[\Delta^n(\delta)]_n + \langle\Delta^n(\delta)\rangle_n\} \\ &\leqslant 3\delta^2\{[\Delta^n(\delta)]_n + \langle\Delta^n(\delta)\rangle_n\}. \end{aligned} \tag{40}$$

由于 $[\Delta^n(\delta)]$ 和 $\langle\Delta^n(\delta)\rangle$ 相互控制, 故由 (40) 式可见, $(m^n(\delta))^2$ 受序列 $6\delta^2[\Delta^n(\delta)]$ 和 $6\delta^2\langle\Delta^n(\delta)\rangle$ 控制.

因此, 如果满足条件 (C), 则对于充分小的 δ (例如, $\delta^2 < \min(\varepsilon, b(\sigma_1^2+1)/6)$), 有

$$\overline{\lim_n}\mathbf{P}\{6\delta^2\langle\Delta^n(\delta)\rangle_n > b\} = 0,$$

从而, 由 §3 定理 4 的系知 (39) 式成立. 而假如满足条件 (C*), 则对于同一 δ, 有

$$\overline{\lim_n}\mathbf{P}\{6\delta^2[\Delta^n(\delta)]_n > b\} = 0. \tag{41}$$

仍然由 §3 定理 4 的系知, 由于 $|\Delta[\Delta^n(\delta)]_k| \leqslant (2\delta)^2$, 故由 (41) 式得 (39) 式. □

6. 定理 3 的证明 考虑林德伯格条件 (L), 可以通过直接计算验证条件 (C) 和 (1), 以及条件 (C*) 和 (3) 的等价性 (练习题 6).

7. 定理 4 的证明 由林德伯格条件 (L), 可见条件 (A) 成立. 关于条件 (B), 只需注意到, 在 ξ^n 构成鞅 - 差的情形下, 典型分解式 (9) 中的随机变量 B_t^n 可以表为如下形式:

$$B_t^n = -\sum_{k=1}^{[nt]}\int_{\{|x|>1\}} x\mathrm{d}\nu_k^n.$$

所以, 由林德伯格条件 (L), 可见 $B_t^n \xrightarrow{\mathbf{P}} 0$.

8. 平方可积鞅 - 差的情形 这一节的基本定理 —— 定理 1, 是在被加项之和一致渐近小的条件下证明的. 自然提出的问题是, 在没有一致渐近小条件的情况下, 中心极限定理成立的条件如何? 对于独立随机变量的情形, 第三章 §5 的定理 1 就是这样的例子 (其条件是: 二阶矩有限).

我们现在 (不加证明地) 给出这一定理的类似, 但是局限于序列 $\xi^n = (\xi_{nk}, \mathscr{F}_k^n)$, $1 \leqslant k \leqslant n$, 是平方可积鞅 - 差的情形 ($\mathbf{E}(\xi_{nk}^2 < \infty, \mathbf{E}(\xi_{nk}|\mathscr{F}_{k-1}^n) = 0)$.

记 $F_{nk}(x) = \mathbf{P}\{\xi_{nk} \leqslant x|\mathscr{F}_{k-1}^n\}$ 是 ξ_{nk} 关于 $\mathscr{F}_{k-1}^n$ 的正则分布函数. 设 $\Delta_{nk} = \mathbf{E}(\xi_{nk}^2|\mathscr{F}_{k-1}^n)$.

定理 5 如果对于平方可积鞅 - 差 $\xi^n = (\xi_{nk}, \mathscr{F}_k^n)$, $0 \leqslant k \leqslant n$, $n \geqslant 1$, 满足条件:

$$\sum_{k=1}^{[nt]}\Delta_{nk} \xrightarrow{\mathbf{P}} \sigma_t^2, \quad 0 \leqslant \sigma_t^2 < \infty, \quad 0 \leqslant t \leqslant 1,$$

而对于任意 $\varepsilon > 0$,

$$\sum_{k=1}^{[nt]} \int_{\{|x|>\varepsilon\}} |x| \left| F_{nk}(x) - \Phi\left(\frac{x}{\sqrt{\Delta_{nk}}}\right) \right| \mathrm{d}x \xrightarrow{\mathbf{P}} 0,$$

且

$$X_t^n \xrightarrow{d} N(0, \sigma_t^2).$$

9. 练习题

1. 设 $\xi_n = \eta_n + \zeta_n, n \geqslant 1$, 其中 $\eta_n \xrightarrow{d} \eta$, 而 $\zeta_n \xrightarrow{d} 0$. 证明 $\xi_n \xrightarrow{d} \eta$.

2. 设 $(\xi(\varepsilon)), n \geqslant 1, \varepsilon > 0$, 是随机变量族, 且对于每一个 $\varepsilon > 0$, 当 $n \to \infty$ 时, 有 $\xi_n \xrightarrow{\mathbf{P}} 0$. 例如, 利用第二章 §10 练习题 11 的论断证明, 存在序列 $\varepsilon_n \downarrow 0$, 使 $\xi_n(\varepsilon_n) \xrightarrow{\mathbf{P}} 0$.

3. 设 $(\alpha_k^n), 1 \leqslant k \leqslant n, n \geqslant 1$, 是复数值随机变量, ($\mathbf{P}$−a.c.)

$$\sum_{k=1}^{n} |\alpha_k^n| \leqslant C, \quad |\alpha_k^n| \leqslant a_n \downarrow 0.$$

证明 ($\mathbf{P}$−a.c.)

$$\lim_n \prod_{k=1}^{n} (1 + \alpha_k^n) \mathrm{e}^{-\alpha_k^n} = 1.$$

4. 证明定理 1 的注 2 中提出的命题.

5. 证明引理的注中提出的命题.

6. 证明定理 3.

7. 证明定理 5.

§9. 伊藤公式的离散版本

1. 引言 在布朗运动、以及与之同类的过程 (鞅、局部鞅、下鞅 ……) 的随机分析中, *伊藤清* (Itô Kiyosi) *变量替换公式*有特别的作用. 在这一节考虑该公式的*离散型变式*, 并且说明如何经极限过渡, 可以得到*布朗运动的伊藤清公式*.

2. 二次协方差的积分表示 设 $X = (X_n)_{0 \leqslant n \leqslant N}$ 和 $Y = (Y_n)_{0 \leqslant n \leqslant N}$ 是概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$ 上的两个随机变量序列, 且 $X_0 = Y_0 = 0$ 而

$$[X, Y] = ([X, Y]_n)_{0 \leqslant n \leqslant N},$$

其中

$$[X, Y]_n = \sum_{k=1}^{n} \Delta X_k \Delta Y_k \tag{1}$$

是序列 X 和 Y 的二次协方差 (见 §1).

假设给定一绝对连续函数 $F = F(x)$:

$$F(x) = F(0) + \int_0^x f(y)\mathrm{d}y, \tag{2}$$

其中 $f = f(y)$ 是 $\mathbb{R}$ 上的博雷尔函数, 对于任意 $c > 0$, 满足条件

$$\int_{|y|\leqslant c} |f(y)|\mathrm{d}y < \infty.$$

下面提到的变量替换公式, 用序列 $X = (X_n)_{0\leqslant n\leqslant N}$ “自然” 泛函的语言, 给出序列

$$F(X) = (F(X_n))_{0\leqslant n\leqslant N}. \tag{3}$$

表示的概念.

考虑序列 X 和 $f(X) = (f(X_n))_{0\leqslant n\leqslant N}$ 的二次协方差 $[X, f(X)]$, 其中 $f = (x), x \in \mathbb{R}$, 是 (2) 式中的函数. 根据 (1) 式

$$[X, f(X)]_n = \sum_{k=1}^{n} \Delta f(X_k)\Delta X_k = \sum_{k=1}^{n}[f(X_k) - f(X_{k-1})](X_k - X_{k-1}). \tag{4}$$

如果引进两个 “离散积分” (对照 §1 定义 5)

$$I_n(X, f(X)) = \sum_{k=1}^{n} f(X_{k-1})\Delta X_k, \quad 1 \leqslant n \leqslant N, \tag{5}$$

和

$$\widetilde{I}_n(X, f(X)) = \sum_{k=1}^{n} f(X_k)\Delta X_k, \quad 1 \leqslant n \leqslant N, \tag{6}$$

则二次协方差可以表示为

$$[X, f(X)]_n = \widetilde{I}_n(X, f(X)) - I_n(X, f(X)). \tag{7}$$

(当 $n = 0$ 时, 设 $I_0 = \widetilde{I}_0 = 0$).

在我们所考虑的情形下, 对于固定的实数 N, 引进新的 (“逆向”) 序列 $\widetilde{X} = (\widetilde{X}_n)_{0\leqslant n\leqslant N}$, 得

$$\widetilde{X}_n = X_{N-n}. \tag{8}$$

显然,

$$\widetilde{I}_N(X, f(X)) = -I_N(\widetilde{X}, f(\widetilde{X})),$$

且类似地

$$\widetilde{I}_n(X, f(X)) = -\{I_N(\widetilde{X}, f(\widetilde{X})) - I_{N-n}(\widetilde{X}, f(\widetilde{X}))\},$$

因此, 由 (7) 式可得

$$[X, f(X)]_N = -\{I_N(\widetilde{X}, f(\widetilde{X})) + I_N(X, f(X))\},$$

且对于 $1 < n < N$, 有

$$\begin{aligned}[X, f(X)]_n &= -\{I_N(\widetilde{X}, f(\widetilde{X})) - I_{N-n}(\widetilde{X}, f(\widetilde{X}))\} - I_n(X, f(X)) \\ &= -\left\{\sum_{k=N-n+1}^{N} f(\widetilde{X}_{k-1})\Delta\widetilde{X}_k + \sum_{k=1}^{n} f(X_{k-1})\Delta X_k\right\}. \end{aligned} \tag{9}$$

注 有益地注意到, 由 (7) 式和 (9) 式给出的二次协方差 $[X, f(X)]_n$ 表达式的不同结构. (7) 式中的 "积分"

$$I_n(X, f(X)) = \sum_{k=1}^{n} f(X_{k-1})\Delta X_k,$$

是这样形成的: 在区间 $[k-1, k]$ ("左" 端点的) 值 $f(X_{k-1})$, 在此整个区间上都乘以 $\Delta X_k = X_k - X_{k-1}$. 然而, "积分" $\widetilde{I}_n(X, f(X))$ 的组成却不同: 是增量 $\Delta X_k = X_k - X_{k-1}$ 乘以区间 $[k-1, k]$ 的 "右" 端点上的值, 即值 X_k.

这样, 可以说,(7) 式既包含 "正向积分 $I_n(X, f(X))$", 又包含 "反向积分 $\widetilde{I}_n(X, f(X))$". 然而, 在 (9) 式中 (对于序列 X 和 $\widetilde{X}$) 所有 "积分" 都是 "正向的".

3. 伊藤公式的离散版本 由于对于每一个函数 $g = g(x)$, 有

$$g(X_{k-1}) + \frac{1}{2}[g(X_k) - g(X_{k-1})] - \frac{1}{2}[g(X_k) + g(X_{k-1})] = 0,$$

则显然

$$\begin{aligned} F(X_n) = F(X_0) &+ \sum_{k=1}^{n} g(X_{k-1})\Delta X_k + \frac{1}{2}[X, g(X)]_n \\ &+ \sum_{k=1}^{n}\left\{[F(X_k) - F(X_{k-1})] - \frac{g(X_{k-1}) + g(X_k)}{2}\Delta X_k\right\}. \end{aligned} \tag{10}$$

特别, 如果 $g(x) = f(x)$, 其中 $f = f(x)$ 是 (2) 式中的函数, 则

$$F(X_n) = F(X_0) + I_n(X, f(X)) + \frac{1}{2}[X, f(X)]_n + R_n(X, f(X)), \tag{11}$$

其中

$$R_n(X, f(X)) = \sum_{k=1}^{n}\int_{X_{k-1}}^{X_k}\left[f(x) - \frac{f(X_{k-1}) + f(X_k)}{2}\right]\mathrm{d}x. \tag{12}$$

由数学分析中熟知的性质, 如果函数 $f''(x)$连续, 则有 "梯形公式":

$$\int_a^b \left[f(x) - \frac{f(a)+f(b)}{2}\right] \mathrm{d}x = \int_a^b (x-a)(x-b)\frac{f''(\xi(x))}{2!}\mathrm{d}x$$

$$= \frac{(b-a)^3}{2}\int_0^1 x(x-1)f''(\xi[a+(b-a)x])\mathrm{d}x$$

$$= \frac{(b-a)^3}{2}f''(\xi[a+(b-a)\overline{x}])\int_0^1 x(x-1)\mathrm{d}x = -\frac{(b-a)^3}{12}f''(\eta),$$

其中 $\xi(x), \overline{x}$ 和 η 是区间 $[a,b]$ "中间" 的点.

因此, 由 (12) 式, 有

$$R_n(X, f(X)) = -\frac{1}{12}\sum_{k=1}^n f''(\eta_k)(\Delta X_k)^3,$$

其中 $X_{k-1} \leqslant \eta_k \leqslant X_k$. 由上面的表达式, 易见

$$|R_n(X, f(X))| \leqslant \frac{1}{12}\sup f''(\eta) \times \sum_{k=1}^n |\Delta X_k|^3, \tag{13}$$

其中 sup 对于满足 $\min(X_0, X_1, \cdots, X_n) \leqslant \eta \leqslant \max(X_0, X_1, \cdots, X_n)$ 的一切 η 值来求:

我们把 (11) 式称做*伊藤公式的离散版本*. 需要强调, 该公式的右侧由如下三部分构成: "离散积分" $I_n(X, f(X))$, 二次协方差 $[X, f(X)]_n$, 和 "余" 项 $R_n(X, f(X))$. "余" 项的名称在于说明, 当由相应的极限过程向连续情形过渡时, $R_n(X, f(X))$ 的极限为 0 (详见第 5 小节).

4. 例

例 1 如果 $f(x) = a + bx$, 则 $R_n(X, f(X)) = 0$, 这时 (11) 式有如下形式:

$$F(X_n) = F(X_0) + I_n(X, f(X)) + \frac{1}{2}[X, f(X)]_n. \tag{14}$$

(对照下面的 (19) 式).

例 2 设

$$f(x) = \operatorname{sign} x = \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0, \end{cases}$$

而 $F(x) = |x|$.

假设 $X_k = S_k$, 其中 $S_k = \xi_1 + \cdots + \xi_k, k \geqslant 1$, 而 $\xi_1, \xi_2, \cdots$ 是独立伯努利随机变量序列, 且每一个 $\xi_k (k \geqslant 1)$ 都以概率 $1/2$ 分别取 ± 1 为值.

如果设 $S_0 = 0$, 则直接由 (11) 式可以求得

$$|S_n| = \sum_{k=1}^n (\operatorname{sign} S_{k-1})\Delta S_k + N_n, \tag{15}$$

其中 $N_n = \#\{0 \leqslant k < n, S_k = 0\}$ 是序列 $S_0, S_1, \cdots$ 中 "0" 的个数.

包含在 (14) 式中的 "离散积分"

$$\left(\sum_{k=1}^{n}(\text{sign}S_{k-1})\Delta S_k\right)_{n\geqslant 1}$$

是鞅. 因此, 由 (15) 式可见

$$\mathbf{E}|S_n| = \mathbf{E}N_n. \tag{16}$$

因为 (练习题 2)

$$\mathbf{E}|S_n| \sim \sqrt{\frac{2n}{\pi}}, \quad n \to \infty, \tag{17}$$

所以由 (16) 式可见

$$\mathbf{E}N_n \sim \sqrt{\frac{2n}{\pi}}, \quad n \to \infty. \tag{18}$$

换句话说, 在随机游动 $S_0, S_1, \cdots, S_n$ 中 "平局" 的平均次数按量级 $\sqrt{n}$ 增长, 而不是像最初感觉的那样好像 "按量级 n 增长". 需要指出, 性质 (18) 十分直接地与反正弦定律 (见第一章 §10) 相联系, 并且实际上可以由反正弦定律得到.

5. 布朗运动的伊藤变量替换公式 设 $B = (B_t)_{0\leqslant t\leqslant 1}$ 是标准 ($B_0 = 0, \mathbf{E}B_t = 0, \mathbf{E}B_t^2 = t$) 布朗运动 (见第二章 §13), 而 $X_k = B_{k/n}, k = 0, 1, \cdots, n$.

运用 (11) 式, 可得如下结果

$$\begin{aligned} F(B_1) = F(B_0) &+ \sum_{k=1}^{n} f(B_{(k-1)/2})\Delta B_{k/n} \\ &+ \frac{1}{2}[f(B_{\cdot/n}), B_{\cdot/n}]_n + R_n(B_{\cdot/n}, f(B_{\cdot/n})). \end{aligned} \tag{19}$$

由布朗运动的理论 (例如, 可参见 [11], [17], [77]), 知

$$\sum_{k=1}^{n}|B_{k/n} - B_{(k-1)/n}|^3 \xrightarrow{\mathbf{P}} 0, \quad n \to 0. \tag{20}$$

从而, 如果函数 $f = f(x)$ 有二阶导数, 且对于某个常数 $C > 0, f''(x) \leqslant C, x \in \mathbb{R}$, 则由估计 (13) 式, 可见 $R_n(B_{\cdot/n}, f(B_{\cdot/n})) \xrightarrow{\mathbf{P}} 0$.

仍然由布朗运动的理论知, 对于任意博雷尔函数 $f = f(x) \in L^2_{\text{loc}}$ (即对于任意常数 $C > 0$, 满足

$$\int_{|x|\leqslant C} f^2(x)dx < \infty$$

的函数), "离散积分"

$$\sum_{k=1}^{n} f(B_{(k-1)/n})\Delta B_{k/n}$$

(在依概率收敛的意义上显然) 存在极限, 记作 $\int_0^1 f(B_s)\mathrm{d}B_s$, 并且称做布朗运动的伊藤清随机积分.

这样, 由 (19) 式可见, 该式中的 "余" 项 $R_n(B_{\cdot/n}, f(B_{\cdot/n})) \xrightarrow{\mathbf{P}} 0$, "离散积分"

$$\sum_{k=1}^{n} f(B_{(k-1)/n})\Delta B_{k/n}$$

(依概率) 收敛于 "随机积分" $\int_0^1 f(B_s)\mathrm{d}B_s$, 从而二次协方差

$$[B_{\cdot/n}, f(B_{\cdot/n})] \quad (= [f(B_{\cdot/n}), B_{\cdot/n}])$$

依概率收敛的极限, 自然应当记为

$$[B, f(B)]_1.$$

这样, 如果函数 $f = f(x)$ 由二阶导数, 且对于某个常数 $C > 0, |f''(x)| \leqslant C, x \in \mathbb{R}$, 且 $f \in L^2_{\mathrm{loc}}$, 则有如下公式:

$$F(B_1) = F(0) + \int_0^1 f(B_s)\mathrm{d}B_s + \frac{1}{2}[B, f(B)]_1. \tag{21}$$

这时,

$$[B, f(B)]_1 = \int_0^1 f'(B_s)\mathrm{d}s. \tag{22}$$

于是,

$$F(B_1) = F(0) + \int_0^1 f(B_s)\mathrm{d}B_s + \frac{1}{2}\int_0^1 f'(B_s)\mathrm{d}s, \tag{23}$$

或者更标准的形式为:

$$F(B_1) = F(0) + \int_0^1 F'(B_s)\mathrm{d}B_s + \frac{1}{2}\int_0^1 F''(B_s)\mathrm{d}s, \tag{24}$$

正是这一公式 (对于 $F \in C^2$) 称做布朗运动的伊藤变量替换公式.

6. 练习题

1. 证明公式 (15).
2. 证明渐近式 (17).
3. 证明公式 (22).
4. 试证明对于任意函数 $F \in C^2$ (24) 式成立.

§10. 保险中破产概率的计算. 鞅方法

1. 破产概率 这一节将要介绍的内容可以很好地演示, 鞅方法如何 (例如为保险公司) 提供破产概率的简单估计.

设 $X=(X_t)_{t\geqslant 0}$ 是描绘所考察保险公司资本演变的随机过程. 以 $X_0=u>0$ 的值表示公司的初始资本. 假设保费收入由公司以固定的速度 $c>0$ (即经时间 Δt, 收入为 $c\Delta t$) 连续地积累. 假设偿付保费的诉求出现在随机的时间 $T_1, T_2, \cdots (0<T_1<T_2<\cdots)$, 而赔偿的相应金额表现为非负随机变量 $\xi_1, \xi_2, \cdots$

由以上的说明可见, 在时间 $t>0$ 保险公司的资本为

$$X_t = u + ct - S_t, \tag{1}$$

其中

$$S_t = \sum_{i\geqslant 1} \xi_i I(T_i \leqslant t). \tag{2}$$

设

$$T = \inf\{t \geqslant 0 : X_t \leqslant 0\}$$

是保险公司的资本首次变为 0 或负值的时间.

如果对于一切 $t\geqslant 0, X_t>0$, 则设 T 等于 $+\infty$. 由完全明显的理由, 自然把时间 T 称为公司的 "破产时间". 我们在下面的主要兴趣是, 求 (或者估计) *破产的概率* $\mathbf{P}\{T<\infty\}$, 或对于任意 $t>0$, *在时刻 t 前破产的概率* $\mathbf{P}\{T\leqslant t\}$.

2. 克拉默 – 伦德伯格模型 求这些破产概率是相当不容易的问题. 不过, 对于*克拉默 – 伦德伯格* (H. Cramér – G. A. Lundeberg) *模型*, 该问题有 (部分) 解. 而所谓克拉默 – 伦德伯格模型的特点是, 它满足如下条件.

A. 假设时间 $\sigma_i = T_i - T_{i-1}, i\geqslant 1 (T_0=0)$, 是独立随机变量, 并且服从指数分布: $\mathbf{P}\{\sigma_i>t\} = \lambda \mathrm{e}^{-\lambda t}, t\geqslant 0, i\geqslant 1$. (见第二章 §3 表 3).

B. 随机变量 $\xi_1, \xi_2, \cdots$ 是独立同分布, 其分布函数 $F(x)=\mathbf{P}\{\xi_1\leqslant x\}$ 满足条件: $F(0)=0$,

$$\mu = \int_0^\infty x \mathrm{d}F(x) < \infty.$$

C. 序列 $(T_1, T_2, \cdots)$ 和 $(\xi_1, \xi_2, \cdots)$ 是独立序列 (按第二章 §5 定义 6 的含义).

设

$$N_t = \sum_{i\geqslant 1} I(T_i \leqslant t), \quad t>0, \tag{3}$$

是描绘 t 时前 (包括时刻 t), 偿付保费诉求数量的过程, 且 $N_0=0$.

因为对于 $k\geqslant 1$,

$$\{T_k > t\} = \{\sigma_1 + \cdots + \sigma_k > t\} = \{N_t < k\},$$

所以注意到条件 **A**, 且根据第二章 §8 练习题 6, 有

$$\mathbf{P}\{N_t<k\} = \mathbf{P}\{\sigma_1+\cdots+\sigma_k>t\} = \sum_{i=0}^{k-1} \mathrm{e}^{-\lambda t}\frac{(\lambda t)^i}{i!}.$$

从而

$$\mathbf{P}\{N_t = k\} = \mathrm{e}^{-\lambda t}\frac{(\lambda t)^k}{k!} \quad k = 0, 1, \cdots \tag{4}$$

这样, 随机变量 N_t 有泊松分布 (参见第二章 §3 表2), 参数为 λt, 恰好等于数学期望 $\mathbf{E}N_t$.

按 (3) 式构造的过程 $N = (N_t)_{t \geqslant 0}$, 特别更新过程的情况 (第二章 §9 第 4 小节), 称做*泊松过程*. 该过程的轨道是间断的 (确切地说, 是阶梯函数, 右连续且跃度皆等于 1). 与轨道是连续函数的布朗运动 (第二章 §13) 一样, 泊松过程在随机过程论中重要作用. 具体地说, 利用这两个过程, 可以建立具有相当复杂概率结构的随机过程.(独立增量过程就是这种情形的典型例子: 例如. 见 [13], [11], [76]).

3. 克拉默 – 伦德伯格模型下的破产概率 由条件 **C** 可见,

$$\begin{aligned}\mathbf{E}(X_t - X_0) &= ct - \mathbf{E}S_t = ct - \mathbf{E}\sum_i \xi_i I(T_i \leqslant t) = ct - \sum_i \mathbf{E}\xi_i I(T_i \leqslant t) \\ &= ct - \sum_i \mathbf{E}\xi_i \mathbf{E}I(T_i \leqslant t) = ct - \mu\sum_i \mathbf{P}\{T_i \leqslant t\} \\ &= ct - \mu\sum_i \mathbf{P}\{N_i \geqslant i\} = ct - \mu\mathbf{E}N_t = t(c - \lambda\mu).\end{aligned}$$

由此清楚地看到, 公司经营赢利 (即 $\mathbf{E}(X_t - X_0) > 0$) 的条件, 表现为

$$c > \lambda\mu. \tag{5}$$

在下面的分析中, 起重要作用的下面的函数

$$h(z) = \int_0^\infty (\mathrm{e}^{zx} - 1)\mathrm{d}F(x), \quad z \geqslant 0, \tag{6}$$

其中 $h(z) = \widehat{F}(-z) - 1$, 而

$$\widehat{F}(s) = \int_0^\infty \mathrm{e}^{-sx}\mathrm{d}F(x)$$

是拉普拉斯 – 斯蒂尔切斯 (P. S. Laplace – T. J. Stieltjes) 变换 (s 是复数).

设

$$g(z) = \lambda h(z) - cz, \quad \xi_0 = 0,$$

则对于任何使 $h(r) < \infty$ 的 $r > 0$, 有

$$\begin{aligned}\mathbf{E}\mathrm{e}^{-r(X_t - X_0)} &= \mathbf{E}\mathrm{e}^{-r(X_t - \mu)} = \mathrm{e}^{-rct}\mathbf{E}\mathrm{e}^{r\sum\limits_{i=0}^{N_t}\xi_i} = \mathrm{e}^{-rct}\sum_{n=0}^{\infty}\mathbf{E}\mathrm{e}^{r\sum\limits_{i=0}^{N_t}\xi_i}\mathbf{P}\{N_t = n\} \\ &= \mathrm{e}^{-rct}\sum_{n=0}^{\infty}[1 + h(r)]^n\frac{\mathrm{e}^{-\lambda t}(\lambda t)^n}{n!} = \mathrm{e}^{-rct}\mathrm{e}^{\lambda t h(r)} = \mathrm{e}^{t[\lambda h(r) - cr]} = \mathrm{e}^{tg(r)}.\end{aligned}$$

对于 $s<t$, 类似地可以得到

$$\mathbf{E}\mathrm{e}^{-r(X_t-X_s)}=\mathrm{e}^{(t-s)g(r)}. \tag{7}$$

设 $\mathscr{F}_0^X=\sigma(X_s,s\leqslant t)$. 由于过程 $X=(X_t)_{t\geqslant 0}$ 是独立增量过程 (练习题 2), 则 ($\mathbf{P}$ - a.c.)

$$\mathbf{E}(\mathrm{e}^{-r(X_t-X_s)}|\mathscr{F}_s^X)=\mathbf{E}\mathrm{e}^{-r(X_t-X_s)}=\mathrm{e}^{(t-s)g(r)},$$

从而

$$\mathbf{E}(\mathrm{e}^{-rX_t-tg(r)}|\mathscr{F}_s^X)=\mathrm{e}^{-rX_s-sg(r)}, \tag{8}$$

设

$$Z_t=\mathrm{e}^{-rX_s-sg(r)},\quad t\geqslant 0. \tag{9}$$

由 (8) 式可见, 显然

$$\mathbf{E}(Z_t|\mathscr{F}_s^X)=Z_s,\quad s\leqslant t. \tag{10}$$

与 §1 定义 1 类似, 自然称过程 $Z=(Z_t)_{t\geqslant 0}$ (关于 $\sigma-$ 代数 "流" $(\mathscr{F}_t^X)_{t\geqslant 0}$) 是鞅. 注意, 对于所考虑的情形, $\mathbf{E}|Z_t|<\infty, t\geqslant 0$ (对照 §1 的性质 1). 与 §1 定义 3 类似, 值域为 $[0,+\infty]$ 的随机变量 $\tau=\tau(\omega)$ 是马尔可夫时间, 或 (关于 $\sigma-$ 代数 "流" $(\mathscr{F}_t^X)_{t\geqslant 0}$) 不依赖于将来的随机变量, 假如对于每一个 $t\geqslant 0$, 有

$$\{\tau(\omega)\leqslant t\}\in\mathscr{F}_t^X.$$

对于现在考虑的连续时间的情形, §2 的定理 1 (在明显地改变记号的情形下) 仍然成立. 特别, 对于马尔可夫时间 τ, 有

$$\mathbf{E}Z_{t\wedge\tau}=\mathbf{E}Z_0. \tag{11}$$

设 $T=\tau$. 那么, 由 (9) 式和 (11) 式, 可见对于任意 $t>0$, 有

$$\begin{aligned}\mathrm{e}^{-ru}&=\mathbf{E}\mathrm{e}^{-rX_{t\wedge T}-(t\wedge T)g(r)}\geqslant\mathbf{E}[\mathrm{e}^{-rX_{t\wedge T}-(t\wedge T)g(r)}|T\leqslant t]\mathbf{P}\{T\leqslant t\}\\&=\mathbf{E}[\mathrm{e}^{-rX_T-Tg(r)}|T\leqslant t]\mathbf{P}\{T\leqslant t\}\geqslant\mathbf{E}[\mathrm{e}^{-Tg(r)}|T\leqslant t]\mathbf{P}\{T\leqslant t\}\\&\geqslant\min_{0\leqslant s\leqslant t}\mathrm{e}^{-sg(r)}\mathbf{P}\{T\leqslant t\}.\end{aligned}$$

从而

$$\mathbf{P}\{T\leqslant t\}\leqslant\frac{\mathrm{e}^{-ru}}{\min\limits_{0\leqslant s\leqslant t}\mathrm{e}^{-sg(r)}}=\mathrm{e}^{-ru}\max_{0\leqslant s\leqslant t}\mathrm{e}^{sg(r)}. \tag{12}$$

我们现在更进一步考虑函数

$$g(r)=\lambda h(r)-cr.$$

显然, $g(0)=0$, (由于 (5) 式) $g'(0)=\lambda\mu-c<0$, 且 $g''(r)=\lambda h''(r)\geqslant 0$. 因此存在唯一正值 $r=R$, 使 $g(R)=0$.

注意, 对于 $r>0$,

$$\int_0^\infty \mathrm{e}^{rx}[1-F(x)]\mathrm{d}x=\int_0^\infty\int_x^\infty \mathrm{e}^{rx}\mathrm{d}F(y)\mathrm{d}x=\int_0^\infty\left(\int_0^y \mathrm{e}^{rx}\mathrm{d}x\right)\mathrm{d}F(y)$$
$$=\frac{1}{r}\int_0^\infty(\mathrm{e}^{ry}-1)\mathrm{d}F(y)=\frac{1}{r}h(r).$$

由此且由等式 $\lambda h(R)-cR=0$ 可见, R 的值是方程

$$\frac{\lambda}{c}\int_0^\infty \mathrm{e}^{rx}[1-F(x)]\mathrm{d}x=1 \tag{13}$$

(且在这种情况下是唯一) 的根.

现在, 如果在 (12) 式中设 $r=R$, 则对于每一个 $t>0$, 有

$$\mathbf{P}\{T\leqslant t\}\leqslant \mathrm{e}^{-Ru}, \tag{14}$$

从而

$$\mathbf{P}\{T<\infty\}\leqslant \mathrm{e}^{-Ru}, \tag{15}$$

于是, 证明了下面的定理.

定理 *假设对于克拉默 – 伦德伯格模型, 满足如下条件 A,B,C 和 $\lambda\mu<c$. 那么, 破产概率 $\mathbf{P}\{T\leqslant t\}$ 和 $\mathbf{P}\{T<\infty\}$ 满足不等式 (14) 和 (15), 其中 R 是方程 (13) 的 (唯一) 正根.*

4. 连续时间的情形 在上面进行的证明中用到 (11) 式, 如曾经指出的那样, 对于连续情形, (11) 式的正确性, 可以由 §2 定理 1 相应的类似得到 (§2 定理 1, 是关于在把时间换为随机马尔可夫时间时鞅性的不变性).(这一结果的证明, 例如, 参见专著 [41] 的 §3.2). 不过, 假如将 (条件 **A** 中的) 变量 $\sigma_i, i=1,2,\cdots$, "服从指数分布", 换成 "服从 (离散型) 几何分布: $\mathbf{P}\{\sigma_i=k\}=q^{k-1}p, k\geqslant 1$", 那么只需引用 §2 中证明的定理 1.

以上引进的要求借鉴连续时间随机过程理论的结果, 是有益的至少在于, 在应用中出现的模型, 并不是在离散时间而是在连续时间起作用的模型.

5. 练习题

1. 证明过程 $N=(N_t)_{t\geqslant 0}$ (在条件 **A** 下) 是独立增量过程.
2. 证明过程 $X=(X_t)_{t\geqslant 0}$ 也是独立增量过程.
3. 考虑克拉默 – 伦德伯格模型, 并且假设变量 $\sigma_i, i=1,2,\cdots$, 独立, 并且 "服从几何分布: $\mathbf{P}\{\sigma_i=k\}=q^{k-1}p, k\geqslant 1$". 试表述本节的定理.

§11. 随机金融数学的基本定理. 无仲裁的鞅特征

1. 引言 上一节介绍了鞅论用于证明保险理论的基本定理之一 —— *伦德伯格 - 克拉默定理*. 这一节再考虑鞅论的一种应用 —— 鞅论用于在不确定条件下运行的 "*无仲裁金融市场*" 问题. 下面引进的定理 1 和定理 2, 在随机金融数学中习惯上称为制裁理论的 "*基本定理*". 这两个定理的重要性在于, 其用*鞅的术语*给出了保障所考察的金融市场无仲裁性的条件 (其含义将在下面说明), 以及保障达到确定的金融目标的条件. (关于金融数学, 详见 [100]).

2. 金融数学的某些概念 这里, 将要给出几个必要的定义.

对于讨论的所有问题, 都假设给定某个*固定的概率空间* $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 0}, \mathbf{P})$, 它在价格、金融指数以及金融市场的其他指标的演变中, 是描绘*随机不确定性*的基础. 这里, 我们把 $\mathscr{F}_n$ 中的全体事件视为在时间 n (包括 n) 得到的 "信息". 例如, $\mathscr{F}_n$ 可能包含关于一定金融证卷的价格、金融指数等数值的 "信息".

与 "基本定理" 有关的, 基本对象是 $(B, S)-$ 市场的概念, 其定义如下.

设 $B = (B_n)_{n\geqslant 0}$ 和 $S = (S_n)_{n\geqslant 0}$ 是正值随机变量序列, 其中假设对于每个 $n \geqslant 0$, 随机变量 B_n 是 $\mathscr{F}_{n-1}-$ 可测的 ($\mathscr{F}_{-1} = \mathscr{F}_0$), 而变量 S_n 是 $\mathscr{F}_n-$ 可测的. 为简便计, 在下面的讨论中, 假设 $\sigma-$ 代数 $\mathscr{F}_0$ 是平凡的, 即 $\mathscr{F}_0 = (\varnothing, \Omega)$ (见第二章 §2). 这样, B_0 和 S_0 都是常数. 于是, 按 §1 的术语, 两个序列 $B = (B_n)_{n\geqslant 0}$ 和 $S = (S_n)_{n\geqslant 0}$ 都是*随机子序列*, 并且序列 $B = (B_n)_{n\geqslant 0}$ 同时又是*可预测的* (因为 B_n 是 $\mathscr{F}_{n-1}-$ 可测的).

按其金融意义, 序列 $B = (B_n)_{n\geqslant 0}$ 是描绘*银行账户* ("bank account", "meney account")"单位" 演变的序列. 这时, 随机变量 B_n 是 $\mathscr{F}_{n-1}-$ 可测性表示, 在时间 n (例如, 今天) 银行账户的值, 在时间 $n-1$ (例如, 昨天) 已经是完全已知的.

如果对于 $n \geqslant 1$, 记

$$r_n = \frac{\Delta B_n}{B_{n-1}}, \tag{1}$$

其中 $\Delta B_n = B_n - B_{n-1}$, 则显然 B_n 可以表示为

$$B_n = (1 + r_n)B_{n+1}, \quad n \geqslant 1, \tag{2}$$

其中随机变量 r_n 是 $\mathscr{F}_{n-1}-$ 可测的, 并且 $r_n > -1$ (因为根据条件 $B_n > 0$). 在金融学文献中, 量 r_n 称做 (*银行*) *利率*.

序列 $S = (S_n)_{n\geqslant 0}$ 与序列 $B = (B_n)_{n\geqslant 0}$ 的差别在于, S_n 是 $\mathscr{F}_n-$ 可测的, 而 B_n 是 $\mathscr{F}_{n-1}-$ 可测的. 实际情况正是这样, 例如, *股票* ("stock", "stocks") *价格*: 对于股票在时间 n 的真实价格, 只有在其*报价时* (即 "今天", 像银行账户那样而不是 "昨天").

类似于 (银行) 利率, 对于股票 $S = (S_n)_{n\geqslant 0}$, 可以引进所谓 "*市场*" *利率*:

$$\rho_n = \frac{\Delta S_n}{S_{n-1}}, \quad n \geqslant 1. \tag{3}$$

由此明显可见, 对于一切 $\rho_n > -1$, 有

$$S_n = (1+\rho_n)S_{n-1}, \tag{4}$$

因为 (根据假设) 所有 $S_n > 0$.

由 (2) 式和 (4) 式, 可见

$$B_n = B_0 \prod_{k=1}^{n}(1+r_k), \tag{5}$$

$$S_n = S_0 \prod_{k=1}^{n}(1+\rho_k). \tag{6}$$

在金融学文献中, 通常称这些公式按单利的类型形成. 在许多问题中, 按复利

$$B_n = B_0 \exp\left\{\sum_{k=1}^{n}\widehat{r}_k\right\}, \quad S_n = S_0 \exp\left\{\sum_{k=1}^{n}\widehat{\rho}_k\right\}, \tag{7}$$

的类型形成的概念也很重要, 其中

$$\widehat{r}_n = \ln(1+r_k) = \ln\left(1+\frac{\Delta B_k}{B_{k-1}}\right), \tag{8}$$

$$\widehat{\rho}_n = \ln(1+\rho_k) = \ln\left(1+\frac{\Delta S_k}{S_{k-1}}\right). \tag{9}$$

习惯上把这些变量称为 "对数利润", "归还", "偿还".

上面用描述方法引进的两过程 $B = (B_n)_{n\geqslant 0}$ 和 $S = (S_n)_{n\geqslant 0}$, 按定义构成 $(B,S)-$ 金融市场, 由银行账户 B 和股票 S 两种资产组成.

注 显然, 因为现实金融市场通常包含大量不同性质的资产 (例如, 见 [100]), 这样的 $(B,S)-$ 市场仅仅是现实金融市场最简单的模型. 然而, 就是在这种简单的情形下, 在研究许多纯金融 - 经济来源的问题时, 已经可以仔细研究并且作为例子说明, 鞅论方法的有效性. (例如, 关于在 $(B,S)-$ 市场上无仲裁可能性的问题, 就属于这种情形. 而关于这一问题, 将在下面引进的 "第一基本定理" 中给出答案).

3. 有价证卷总存和自筹资金总存 现在给出有价证券的总存量及其资本的定义, 以及引进自筹资金总存量的重要概念.

设 $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 0}, \mathbf{P})$ 是所考察的过滤概率空间, 其中 $\mathscr{F}_0 = (\varnothing, \Omega)$; 而 $\pi = (\beta, \gamma)$ 是可预测序列偶: $\beta = (\beta_n)_{n\geqslant 0}$和 $\gamma = (\gamma_n)_{n\geqslant 0}$.

除变量 β_n 和 $\gamma_n, n \geqslant 0$ 的可预测性, 即 β_n 和 γ_n 为 $\mathscr{F}_{n-1}-$ 可测的条件之外, 其中 $(\mathscr{F}_{-1} = \mathscr{F}_0)$, 对于 β_n 和 $\gamma_n (n \geqslant 0)$ 的可能值不再加任何限制. 特别, 这些变量可以取分数和负数为值.

随机变量 β_n 和 γ_n 的含义相应为, 在时间 n 银行账户的 "个数" 和股票的 "张数".

我们称在所考察的 $(B,S)-$ 市场上, $\pi=(\beta,\gamma)$ 是有价证券的*总存量*.

与每一总存量 $\pi=(\beta,\gamma)$, 有与其相应的*资本* $X^\pi=(X_n^\pi)_{n\geqslant 0}$ 相联系, 其中

$$X_n^\pi=\beta_n B_n+\gamma_n S_n, \tag{10}$$

把 $\beta_n B_n$ 解释为银行账户上的货币资金, 而 $\gamma_n S_n$ 是时间 n 股票的价值. 序列 β 和 γ 的可预测性也是清楚的:"明天" 有价证卷的总存量, 应该在 "今天" 编制.

特别, 下一个重要的概念 —— "自筹资金" 总量的概念, 反映分析这样 $(B,S)-$ 市场: 在这样的市场上 "无论资金流出, 还是自外部流入" 都不存在. 从形式的观点看, 相应的定义是由下面的方式给出的.

由 "离散微分公式" $\Delta(a_n b_n)=a_n\Delta b_n+b_{n-1}\Delta a_n$, 可见资金的增量 ΔX_n^π $(=X_n^\pi-X_{n-1}^\pi)$ 可以表示为

$$\Delta X_n^\pi=[\beta_n\Delta B_n+\gamma_n\Delta S_n]+[B_{n-1}\Delta\beta_n+S_{n-1}\Delta\gamma_n]. \tag{11}$$

资金的*实际的*变化只与银行账户和股票价格值的 "市场" 变化有关, 即与量 $\beta_n\Delta B_n+\gamma_n\Delta S_n$ 的变化有关. (11) 式右侧的第二项, 即 $B_{n-1}\Delta\beta_n+S_{n-1}\Delta\gamma_n$ 是 $\mathscr{F}_{n-1}-$ 可测变量, 并且在时间 n 变量 X_{n-1}^π 任何增加或减少都不可能. 于是它应该等于 0.

原则上, 资本的可能发生变化, 不仅因为利率 (r_n 和 $\rho_n, n\geqslant 1$) 的 "市场" 变化, 例如还可能因为资本从外部流入, 由于业务费用的支出, 等等.

以后, 这些可能性都不考虑, 假设全部考察的总存 $\pi=(\beta,\gamma)$ 满足条件: 对于一切 $n\geqslant 1$,

$$\Delta X_n^\pi=\beta_n\Delta B_n+\gamma_n\Delta S_n. \tag{12}$$

在随机金融数学中, 通常把这样的总存称为*自筹的* (self-financing).

4. 随机金融数学的第一基本定理 由 (12) 式可见, 对于自筹总存量 $\pi=(\beta,\gamma)$,

$$X_n^\pi=X_0^\pi+\sum_{k=1}^n(\beta_k\Delta B_k+\gamma_k\Delta S_k), \tag{13}$$

且由于

$$\Delta\left(\frac{X_n^\pi}{B_n}\right)=\gamma_n\Delta\left(\frac{S_n}{B_n}\right), \tag{14}$$

则

$$\frac{X_n^\pi}{B_n}=\frac{X_0^\pi}{B_0}+\sum_{k=1}^n\gamma_k\Delta\left(\frac{S_k}{B_k}\right). \tag{15}$$

固定某个 $N\geqslant 1$, 并考察 $(B,S)-$ 市场在时间 $n=0,1,\cdots,N$ 的变化.

定义 1 称在时间 N 的自筹资金总存量 (或自筹资金策略) $\pi=(\beta,\gamma)$, *实施仲裁, 或实施仲裁的可能性*, 如果 $X_0^\pi=0, X_N^\pi\geqslant 0$ ($\mathbf{P}$ - a.c.), 且以大于 0 的 $\mathbf{P}-$ 概率, 有 $X_N^\pi>0$, 即 $\mathbf{P}\{X_N^\pi>0\}>0$.

定义 2 称 (在时间 N) 在 $(B,S)-$ 市场上无仲裁或无仲裁可能性, 如果对于 $X_0^\pi=0$ 和 $\mathbf{P}\{X_N^\pi\geqslant 0\}=1$ 的任何总存量 $\pi=(\beta,\gamma)$, 实际上 $\mathbf{P}\{X_N^\pi=0\}=1$, 即仅以等于 0 的 P - 概率可能使 $X_N^\pi>0$.

直观上, 在无仲裁市场上不可能出现如下情况: 对于某一总存量存在得到无风险收入的可能性.

显然, 关于某一 $(B,S)-$ 市场是否无仲裁, 因而在一定意义上是 "正确的"、"合理的" 解的问题, 与序列 $B=(B_n)_{n\geqslant 0}$ 和 $S=(S_n)_{n\geqslant 0}$ 的概率 - 统计性质有关, 从而与包含在过滤概率空间 $(\Omega,\mathscr{F},(\mathscr{F}_n)_{n\leqslant N},\mathbf{P})$ 上结构中的假设条件有关.

值得注意的是, 鞅论可以相当有效地描绘保障无仲裁可能性的条件. 可以得到的甚至更多. 具体地说, 有下面的定理.

定理 1 ("第一基本定理") 假设过滤概率空间 $(\Omega,\mathscr{F},(\mathscr{F}_n)_{n\leqslant N},\mathbf{P})$ 描绘随机不确定性, 其中 $\mathscr{F}_0=(\varnothing,\Omega),\mathscr{F}_N=\mathscr{F}$.

为使定义在 $(\Omega,\mathscr{F},(\mathscr{F}_n)_{n\leqslant N},\mathbf{P})$ 上的 $(B,S)-$ 市场是无仲裁的, 必要且充分条件是, 在 $(\Omega,\mathscr{F})$ 上存在与测度 $\mathbf{P}$ 等价的这样一个测度 $\widetilde{\mathbf{P}}(\widetilde{\mathbf{P}}\sim\mathbf{P})$, 使贴现序列

$$\frac{S}{B}=\left(\frac{S_n}{B_n}\right)_{n\leqslant N}$$

关于测度 $\widetilde{\mathbf{P}}$ 构成鞅:

$$\widetilde{\mathbf{E}}\left|\frac{S_n}{B_n}\right|<\infty,\quad n\leqslant N,$$

且

$$\widetilde{\mathbf{E}}\left(\frac{S_n}{B_n}\Big|\mathscr{F}_{n-1}\right)=\frac{S_{n-1}}{B_{n-1}},\quad n\leqslant N,$$

其中 $\widetilde{\mathbf{E}}$ 表示关于测度 $\widetilde{\mathbf{P}}$ 的数学期望.

注 1 对于向量过程 $S=(S^1,\cdots,S^d),d<\infty$, 定理的结论仍然成立 (见 [100] 中第 V 章 §2b).

注 2 根据完全明显的原因, 定理中的测度 $\widetilde{\mathbf{P}}$ 通常称做鞅测度.

我们以

$$\mathbf{M}(\mathbf{P})=\left\{\widetilde{\mathbf{P}}\sim\mathbf{P}:\frac{S}{B}\text{ 是 }\widetilde{\mathbf{P}}-\text{ 鞅}\right\}$$

表示测度 $\widetilde{\mathbf{P}}$ 的全体: 测度 $\widetilde{\mathbf{P}}$ 与测度 $\mathbf{P}$ 等价且序列

$$\frac{S}{B}=\left(\frac{S_n}{B_n}\right)_{n\leqslant N}$$

关于测度 $\widetilde{\mathbf{P}}$ 是鞅.

以记号 **NA** 表示无仲裁 (No Arbitrage).

那么, 定理 1 的结论可以表示为:

$$\mathbf{NA}\Leftrightarrow\mathbf{M}(\mathbf{P})\neq\varnothing. \tag{16}$$

证明 *充分性* 设 $\widetilde{\mathbf{P}}$ 是 $\mathbf{M}(\mathbf{P})$ 中的鞅测度, $\pi=(\beta,\gamma)$ 是总存量, 且 $X_0^\pi=\beta_0B_0+\gamma_0S_0=0$. 由 (15) 式可见, 对于 $1\leqslant n\leqslant N$, 有

$$\frac{X_n^\pi}{B_n}=\sum_{k=1}^{n}\gamma_k\Delta\left(\frac{S_k}{B_k}\right). \tag{17}$$

序列

$$\frac{S}{B}=\left(\frac{S_n}{B_n}\right)_{n\leqslant N}$$

关于测度 $\widetilde{\mathbf{P}}$ 是鞅. 因此, 序列 $G=(G_n^\pi)_{0\leqslant n\leqslant N}(G_0^\pi=0)$, 和

$$G_n^\pi=\sum_{k=1}^{n}\gamma_k\Delta\left(\frac{S_k}{B_k}\right),\quad 1\leqslant n\leqslant N,$$

是鞅变换. 因此序列 $(X_n^\pi/B_n)_{0\leqslant n\leqslant N}$ 也是鞅变换.

在仲裁测试或无仲裁的情况下, 不仅需要考虑 $X_0^\pi=0$ 的总存量 π 的情形, 而且需要考虑 $X_0^\pi\geqslant 0$ ($\mathbf{P}$–a.c.) 的总存量 π 的情形. 由于 $\widetilde{\mathbf{P}}\sim\mathbf{P}$, 且 $B_N>0$ ($\mathbf{P}$ – 和 $\widetilde{\mathbf{P}}$–a.c.), 可得

$$\widetilde{\mathbf{P}}\left\{\frac{X_N^\pi}{B_N}\geqslant 0\right\}=1.$$

那么, 将 §1 的定理 3 用于鞅变换 $(X_n^\pi/B_n)_{0\leqslant n\leqslant N}$ 可见, 该序列实际上关于测度 $\widetilde{\mathbf{P}}$ 是鞅. 从而

$$\widetilde{\mathbf{E}}\frac{X_N^\pi}{B_N}=\widetilde{\mathbf{E}}\frac{X_0^\pi}{B_0}=0,$$

而由于

$$\widetilde{\mathbf{P}}\left\{\frac{X_N^\pi}{B_N}\geqslant 0\right\}=1,\quad 故\ \widetilde{\mathbf{P}}\left\{\frac{X_N^\pi}{B_N}=0\right\}=1.$$

由此可见, $X_N^\pi=0$ ($\widetilde{\mathbf{P}}$ – 和 $\mathbf{P}$–a.c.). 从而对于任意自筹总存量 π, 其中 $X_0^\pi=0, X_N^\pi\geqslant 0$ ($\mathbf{P}$–a.c.), 实际上 $X_N^\pi=0$ ($\mathbf{P}$–a. c.). 于是, 根据定义 2 无仲裁可能性.

必要性 我们之准备仅对于一阶段模型 (B,S)– 市场进行证明, 即只证明 $N=1$ 的情形. 即便通过这个简单的例子, 就已经清楚地看出证明的好思路: 利用无仲裁, 明显地建立无论是什么样的鞅测度. 我们将基于 (下面将引进的) 埃舍 (Esher) 变换, 建立这样的测度.(关于的一般情形 ($N\geqslant 1$) 的证明, [100] 中第 V 章 §2b).

不失普遍性可认为 $B_0=B_1=1$. 这里, 由无仲裁可能性的假设可见 (练习题 1)

$$\mathbf{P}\{\Delta S_1>0\}>0\quad 和\quad \mathbf{P}\{\Delta S_1<0\}>0. \tag{18}$$

(我们排除了平凡的情形 $\mathbf{P}\{\Delta S_1=0\}=1$).

由此需要证明, 存在满足如下条件的等价鞅测度 $\widetilde{\mathbf{P}}$ 即 $\widetilde{\mathbf{P}}\sim\mathbf{P}$, 且 $\widetilde{\mathbf{E}}|\Delta S_1|<\infty, \widetilde{\mathbf{E}}|\Delta S_1|=0$.

这由下面的引理即可得到. 不过该引理尚有一般概率的应用.

引理 1 设 $(\Omega,\mathscr{F})=(\mathbb{R},\mathscr{B}(\mathbb{R}))$, 而 $X=X(\omega)$ 是由坐标给定的随机变量 $(X(\omega)=\omega)$; $\mathbf{P}$ 是 $(\Omega,\mathscr{F})$ 上的概率测度:

$$\mathbf{P}\{X>0\}>0, \quad 和 \quad \mathbf{P}\{X<0\}>0. \tag{19}$$

那么, 在 $(\Omega,\mathscr{F})$ 上存在概率测度 $\widetilde{\mathbf{P}}\sim\mathbf{P}$, 使对于任意实数 a, 有

$$\widetilde{\mathbf{E}}\mathrm{e}^{aX}<\infty. \tag{20}$$

特别, $\widetilde{\mathbf{E}}|X|<\infty$, 并且

$$\widetilde{\mathbf{E}}X=0. \tag{21}$$

证明 引进测度 $\mathbf{Q}=\mathbf{Q}(\mathrm{d}x)$, 而 $\mathbf{Q}(\mathrm{d}x)=c\mathrm{e}^{-x^2}\mathbf{P}(\mathrm{d}x)$, 其中 $c=(\mathbf{E}\mathrm{e}^{-X^2})^{-1}$ 规范化常数.

对于任意实数 a, 设

$$\varphi(a)=\mathbf{E}_{\mathbf{Q}}\mathrm{e}^{aX}. \tag{22}$$

其中 $\mathbf{E}_{\mathbf{Q}}$ 表示对于测度 $\mathbf{Q}$ 的数学期望.

设

$$Z_a(x)=\frac{\mathrm{e}^{aX}}{\varphi(a)}. \tag{23}$$

由于 $Z_a(x)>0$, 且 $\mathbf{E}_{\mathbf{Q}}Z_a(X)=1$, 则对于任意实数 a, 测度 $\widetilde{\mathbf{P}}_a$:

$$\widetilde{\mathbf{P}}_a(\mathrm{d}x)=Z_a(x)\mathbf{Q}_a(\mathrm{d}x) \tag{24}$$

是概率测度. 显然 $\widetilde{\mathbf{P}}_a\sim\mathbf{Q}\sim\mathbf{P}$.

注 3 变换 $x\sim\!\!\rightarrow \mathrm{e}^{aX}/\varphi(a)$ 常称做埃舍变换. 在下面将要看到, 对于 a 的某个特殊值 a_*, 测度 $\widetilde{\mathbf{P}}=\widetilde{\mathbf{P}}_{a_*}$ 具有 (鞅) 性质 (21). 正是这个测度通常称为埃舍测度 (或埃舍鞅测度).

由于 $\varphi''(a)>0$, 可见对于一切实数 a 定义的函数 $\varphi=\varphi(a)$ 是严格凹 (即向下凸) 的.

设 $\varphi_*=\inf\{\varphi(a):a\in\mathbb{R}\}$. 有两种可能的情形: 1) 存在 a_*, 使 $\varphi(a_*)=\varphi_*$; 2) 这样的有限 a_* 不存在.

对于第一种情形, $\varphi'(a_*)=0$. 因此

$$\mathbf{E}_{\widetilde{\mathbf{P}}_{a_*}}X=\mathbf{E}_{\mathbf{Q}}\frac{X\mathrm{e}^{a_*X}}{\varphi(a_*)}=\frac{\varphi'(a_*)}{\varphi(a_*)}=0,$$

且可以取测度 $\mathbf{P}_{a^*}$ 做所要求的测度 $\widetilde{\mathbf{P}}$.

到现在为止, 我们还没有用到条件 “无仲裁性” (19). 不难看到 (练习题 2), “无仲裁性” 条件排除可能性 2). 这样只剩下可能性 1), 而这种情形已经讨论过.

于是, 当 $N=1$ 时 (存在鞅测度的)“必要性” 得证. 关于的一般情形 $N\geqslant 1$, 我们已经给读者指出, 可以借鉴 [100] 中第 V 章 §2d 的内容. □

5. 例 下面举一些无仲裁 (B, S)– 市场的例.

例 1 假设 (B, S)– 市场由 (5) 式和 (6) 式描绘, 其中 $1 \leqslant k \leqslant N$, 且对于一切 $1 \leqslant k \leqslant N, r_k = r$ (常数); 而 $\rho = (\rho_1, \rho_2, \cdots, \rho_N)$, 是独立同分布、只有 a 和 $b(a < b)$ 两个可能值的 (伯努利) 随机变量序列: $\mathbf{P}\{\rho_1 = a\} = q, \mathbf{P}\{\rho_1 = b\} = p, p + q = 1, 0 < p < 1$. 设, 这时

$$-1 < a < r < b. \tag{25}$$

这样描绘的 (B, S)– 市场称做 CRR – 模型, 其中是模型的作者姓氏的字头: 考克斯 (J. C. Cox), 罗斯 (R. A. Ross), 鲁宾斯坦 (M. Rubinstein), (详见 [100]).

由于在该模型中

$$\frac{S_n}{B_n} = \left(\frac{1+\rho_n}{1+r}\right)\frac{S_{n-1}}{B_{n-1}},$$

则显然鞅测度 $\widetilde{\mathbf{P}}$ 应当满足

$$\widetilde{\mathbf{E}}\frac{1+\rho_n}{1+r} = 1,$$

即应当满足 $\widetilde{\mathbf{E}}\rho_n = r$.

如果记 $\widetilde{p} = \widetilde{\mathbf{P}}\{\rho_n = b\}, \widetilde{q} = \widetilde{\mathbf{P}}\{\rho_n = a\}$, 则对于任意 $n \geqslant 1$, 有

$$\widetilde{p} + \widetilde{q} = 1, \quad b\widetilde{p} + a\widetilde{q} = r.$$

由此, 得

$$\widetilde{p} = \frac{r-a}{b-a}, \quad \widetilde{q} = \frac{b-r}{b-a}. \tag{26}$$

在上面的情形下, 全部 "随机性" 决定于伯努利序列 $\rho = (\rho_1, \rho_2, \cdots, \rho_N)$. 假设 $\Omega = \{a, b\}^N$, 即基本事件空间由序列 $(x_1, \cdots, x_N)$ 构成, 其中 $x_i = a$ 或 b. (关于空间 Ω 构造的这一特别的, (确切一点说) "坐标的" 构造的假设, 并不失研究的普遍性; 关于这一点, 亦见第 6 小节定理 2 中充分性证明的末尾).

作为练习 (练习题 3), 需要证明, 由

$$\widetilde{\mathbf{P}}(x_1, \cdots, x_N) = \widetilde{p}^{\nu_b(x_1, \cdots, x_N)}\widetilde{q}^{N-\nu_b(x_1, \cdots, x_N)}, \tag{27}$$

定义的测度 $\widetilde{\mathbf{P}}(x_1, \cdots, x_N)$ 是鞅测度, 并且是唯一的, 其中

$$\nu_b(x_1, \cdots, x_N) = \sum_{i=1}^{N} I_b(x_i)$$

是等于 b 的 x_i 的个数.

由 (27) 式明显可见 $\widetilde{\mathbf{P}}\{\rho_n = a\} = \widetilde{q}, \mathbf{P}\{\rho_n = b\} = \widetilde{p}$.
这样, 由定理 1 可见, CRR – 模型是无仲裁 (B, S)– 市场的例.

例 2 假设 (B,S)– 市场有如下构造: 对于一切 $n=0,1,\cdots,N$, 有 $B_n=1$, 而

$$S_n = S_0 \exp\left\{\sum_{k=1}^{n} \widehat{\rho}_k\right\}, \quad 1 \leqslant n \leqslant N. \tag{28}$$

设 $\widehat{\rho}=\mu_k+\sigma_k\varepsilon_k$ 为 $\mathscr{F}_{n-1}$– 可测, 其中 $\mu_k>0,\sigma_k>0$, 而 $(\varepsilon_1,\cdots\varepsilon_N)$ 是独立平稳高斯随机变量序列, $\varepsilon_k \sim N(0,1)$.

我们现在利用埃舍条件变换, 在 $(\Omega,\mathscr{F}_N)$ 上建立所要求的鞅测度 $\widetilde{\mathbf{P}}$. 具体地说, 设 $\widetilde{\mathbf{P}}(\mathrm{d}\omega)=Z_N(\omega)\mathbf{P}(\mathrm{d}\omega)$, 其中

$$Z_N(\omega) = \prod_{1\leqslant k\leqslant N} z_k(\omega), \quad 而 \quad z_k(\omega) = \frac{\mathrm{e}^{a_k\widehat{\rho}_k}}{\mathbf{E}(\mathrm{e}^{a_k\widehat{\rho}_k}|\mathscr{F}_{k-1})}; \tag{29}$$

现在, 应当这样选择 $\mathscr{F}_{n-1}$– 可测随机变量 $a_k=a_k(\omega)$, 使得序列 $(S_n)_{0\leqslant n\leqslant N}$ 关于测度 $\widetilde{\mathbf{P}}$ 是鞅, 其中 $\mathscr{F}_0=(\varnothing,\Omega)$.

由于表现 (28), 关于测度 $\widetilde{\mathbf{P}}$ 的鞅性等价于: 对于一切 $1\leqslant n\leqslant N$, (关于原测度 $\mathbf{P}$) 有

$$\mathbf{E}[\mathrm{e}^{(a_n+1)\widehat{\rho}_n}|\mathscr{F}_{n-1}] = \mathbf{E}[\mathrm{e}^{a_n\widehat{\rho}_n}|\mathscr{F}_{n-1}]. \tag{30}$$

由于 $\widehat{\rho}_n=\mu_n+\sigma_n\varepsilon_n$, 故由 (30) 式可见, 应这样选择变量 a_n, 使

$$\mu_n + \frac{\sigma_n^2}{2} = -a_n\sigma_n^2,$$

即

$$a_n = -\frac{\mu_n}{\sigma_n^2} - \frac{1}{2}.$$

对于这样选择变量 $a_n(1\leqslant n\leqslant N)$, 密度 $Z_N(\omega)$ 由下面的公式表示:

$$Z_N(\omega) = \exp\left\{-\sum_{n=1}^{N}\left[\left(\frac{\mu_n}{\sigma_n}+\frac{\sigma_n}{2}\right)\varepsilon_n + \frac{1}{2}\left(\frac{\mu_n}{\sigma_n}+\frac{\sigma_n}{2}\right)^2\right]\right\}. \tag{31}$$

如果一开始对于 $1\leqslant n\leqslant N$, 设 $\mu_n=-\sigma_n^2/2$, 则 $\widetilde{\mathbf{P}}=\mathbf{P}$. 换句话说, 原来的测度 $\mathbf{P}$ 本身就是鞅.

于是, 对于所分析的 (B,S) – 市场 $(B=(B_n)_{0\leqslant n\leqslant N})$ 有: $B_n\equiv 1$, 而 $S=(S_n)_{0\leqslant n\leqslant N}$ 同例 1 一样是无仲裁的, 其中 S_n 由 (28) 式表示. 作为习题 (练习题 4), 要求分析如下问题: 上面建立的鞅测度 $\widetilde{\mathbf{P}}$ 是否唯一.

6. 随机金融数学的第二基本定理 下面将要引进的 (B,S) – 市场完全性的概念, 对于随机金融数学十分重要, 因为 (无论所考察的市场是无仲裁的, 还是有仲裁的) 它与如下很自然的问题有关: 假如对于给定的 $\mathscr{F}_N$– 可测 "自筹资金委托" f_N, 存在自筹资金总存量 π, 使其资本 X_N^π 准确地再现 (或者至少不小于) f_N.

定义 3 称 (B,S)– 市场 (关于时间 N) 为完全的, 或 N– 完全的, 如果任意有界 $\mathscr{F}_N$– 可测 "自筹资金委托" f_N 是可以再现的, 即存在自筹资金总存量 π, 使 $X_N^\pi = f_N$ ($\mathbf{P}$-a.c.).

定理 2 (第二基本定理) 同定理 1 一样, 假设 $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{0\leqslant n\leqslant N}, \mathbf{P})$ 是过滤概率空间, $\mathscr{F}_0 = (\varnothing, \Omega), \mathscr{F}_N = \mathscr{F}$; 在 $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\leqslant N}, \mathbf{P})$ 上给定的 (B,S)– 市场是无仲裁的 ($\mathbf{M}(\mathrm{P}) \neq \varnothing$).

为使此市场是完全的, 必要和充分条件是存在唯一鞅测度 ($|\mathbf{M}(\mathrm{P})| = 1$).

证明 必要性 假设所考虑的市场是完全的, 即对于任意 $\mathscr{F}_N$– 可测有界 "自筹资金委托" f_N, 存在这样的自筹资金总存量 $\pi = (\beta, \gamma)$, 使 $X_N^\pi = f_N$ ($\mathbf{P}$-a.c.). 不失普遍性可以认为 $B_n = 1, 0 \leqslant n \leqslant N$. 因而, 由 (13) 式可见

$$f_N = X_N^\pi = X_0^\pi + \sum_{k=1}^N \gamma_k \Delta S_k. \tag{32}$$

由于所作的无仲裁性假设, 鞅测度的集合 $\mathbf{M}(\mathrm{P}) \neq \varnothing$. 现在证明, 由于完全性假设, 可见鞅测度的唯一性 ($|\mathbf{M}(\mathrm{P})| = 1$).

设 $\mathbf{P}^1$ 和 $\mathbf{P}^2$ 是两个鞅测度, 则关于其中任何一个测度, 序列

$$\left(\sum_{k=1}^n \gamma_k \Delta S_k\right)_{1\leqslant n\leqslant N}$$

是鞅变换.

对于某个集合 $A \in \mathscr{F}_N$, 设 $f_N(\omega) = I_A(\omega)$. 由于 ($\mathbf{P}$–a.c.) 对于某个 π, 有

$$I_A(\omega) = X_N^\pi = X_0^\pi + \sum_{k=1}^N \gamma_k \Delta S_k,$$

则由 §1 定理 3 可见, 序列

$$\left(\sum_{k=1}^n \gamma_k \Delta S_k\right)_{1\leqslant n\leqslant N}$$

是关于每一个测度 $\mathbf{P}^1$ 和 $\mathbf{P}^2$ 鞅. 从而

$$\mathbf{E}_{\mathbf{P}^i} I_A(\omega) = x, \quad i = 1, 2, \tag{33}$$

其中 $\mathbf{E}_{\mathbf{P}^i}$ 表示对测度 $\mathbf{P}^i$ 的数学期望, 而因为 $\mathscr{F}_0 = (\varnothing, \Omega), x = X_0^\pi$ 是常数.

由 (33) 式可见, 对于任意集合 $A \in \mathscr{F}_N$, 有 $\mathbf{P}^1(A) = \mathbf{P}^2(A)$. 从而, 鞅测度的唯一性得证.

充分性 充分性的证明比较复杂, 我们将分为几阶段进行.

1) 考虑无仲裁 (B,S)– 市场 ($\mathbf{M}(\mathrm{P}) \neq \varnothing$), 同时假设鞅测度具有唯一性 ($|\mathbf{M}(\mathrm{P})| = 1$).

注意, 关于鞅测度唯一的假设, 以及关于鞅测度完全性的假设, 都是很强的限制. 此外, 结果表明, 由这些假设自然地可见, 轨道 $S=(S_n)_{0\leqslant n\leqslant N}$ 具有 "条件两点" 的构造. 关于这一点将在下面证明. (具体的例子: 设 $\Delta S_n=\rho_n S_{n-1}$ 为 CRR – 模型, 其中 ρ_n 只有两个可能值, 因此条件概率 $\mathbf{P}(\Delta S_n\in\cdot|\mathscr{F}_{n-1})$ 仅集中在 aS_{n-1} 和 bS_{n-1} 两个点上).

鞅测度的唯一性 ($|\mathbf{M}(\mathrm{P})|=1$) 也是加在滤波 $(\mathscr{F}_n)_{n\leqslant N}$ 构造上的限制. 原来, $\sigma-$ 代数 $\mathscr{F}_n$ 自然应该是价格 $S_0,S_1,\cdots,S_n$ 产生的 $\sigma-$ 代数 (只是现在假设 $B_k\equiv 1,k\leqslant n$). 关于这一点见 [100] 第 610 页的图形, 以及 [100] 第五章的 §4e.

2) 作为证明蕴涵关系 " $|\mathbf{M}(\mathrm{P})|=1\Rightarrow$ 完全性" 路径的中间结果之一, 我们证明如下的重要命题, 它给出了无仲裁市场上完全性的等价特征.

引理 2 为了使无仲裁 $(B,S)-$ 市场是完全的, 必要而且充分的条件是: 在所有鞅测度的集合 $\mathbf{M}(\mathbf{P})$ 上, 存在具有如下性质的测度 $\widetilde{\mathbf{P}}$: 任何有界鞅 $m=(m_n,\mathscr{F}_n,\widetilde{\mathbf{P}})_{0\leqslant n\leqslant N}$, 有 "$(S/B)-$ 表现"

$$m_n=m_0+\sum_{k=1}^{n}\gamma_k^*\Delta\left(\frac{S_k}{B_k}\right),\tag{34}$$

其中 $\gamma_k^*(1\leqslant k\leqslant n)$ 是某一可预测随机变量.

证明引理 2 a) 必要性 设所考察的 $(B,S)-$ 市场是无仲裁的和完全的. (不失普遍性, 可以认为 $B_n=1,0\leqslant n\leqslant N$).

取 $\mathbf{M}(\mathbf{P})$ 中的任意测度 $\widetilde{\mathbf{P}}$, 并且设 $m=(m_n,\mathscr{F}_n,\widetilde{\mathbf{P}})_{0\leqslant n\leqslant N}$ 是某一有限鞅 ($|m_n|\leqslant c,0\leqslant n\leqslant N$). 记 $f_N=m_N$. 那么, 根据完全性的定义 (见定义 3), 存在这样的总存量 $\pi^*=(\beta^*,\gamma^*)$, 使 $X_N^{\pi^*}=f_N$, 且对于一切 $0\leqslant n\leqslant N$, 有

$$X_N^{\pi^*}=x+\sum_{k=1}^{n}\gamma_k^*\Delta S_k,\tag{35}$$

其中 $x=X_0^{\pi^*}$.

由于 $X_N^{\pi^*}=f_N\leqslant c$, 可见序列 $X^{\pi^*}=(X_n^{\pi^*},\mathscr{F}_n,\widetilde{\mathbf{P}})_{0\leqslant n\leqslant N}$ 是鞅 (见 §1 定理 3). 这样, 就有具有同一最终值 $f_N(X_N^{\pi^*}=m_N=f_N)$ 的两个鞅 m 和 X^{π^*}. 但是, 根据鞅性质的定义 $m_n=\mathbf{E}(m_N|\mathscr{F}_n)$ 和 $X_n^{\pi^*}=\mathbf{E}(X_N^{\pi^*}|\mathscr{F}_n),0\leqslant n\leqslant N$. 因此, 列维 (P. P. Lévy) 鞅 m 和 X^{π^*} 相等. 从而, 由 (35) 式知, 对于鞅 $m=(m_n,\mathscr{F}_n,\widetilde{\mathbf{P}})_{0\leqslant n\leqslant N}$, 有 "$S-$表现":

$$m_n=x+\sum_{k=1}^{n}\gamma_k^*\Delta S_k,\quad 1\leqslant n\leqslant N,\tag{36}$$

其中 $x=m_0$.

b) 充分性 现在证明相反的结果 ("$S-$表现" $\Rightarrow$ 完全性).

根据假设的条件, 存在测度 $\widetilde{\mathbf{P}}\in\mathbf{M}(\mathbf{P})$, 使任何有限 $\widetilde{\mathbf{P}}-$ 鞅有 "$S-$ 表现".

作为这样的鞅 $X=(X_n,\mathscr{F}_n,\widetilde{\mathbf{P}})_{0\leqslant n\leqslant N}$, 取 $X_n=\widetilde{\mathbf{E}}(f_N|\mathscr{F}_n)$ 的鞅, 其中 $\widetilde{\mathbf{E}}$ 表示对于测度 $\widetilde{\mathbf{P}}$ 的数学期望, 而 f_N 是定义 3 中提到的 "支付委托", 且对于 f_N 需要求自筹资金总量 $\pi: X_N^\pi=f_N(\widetilde{\mathbf{P}}$ - 和 $\mathbf{P}$-a.c.).

对于有界鞅 $X=(X_n,\mathscr{F}_n,\mathbf{P})_{0\leqslant n\leqslant N}$, 考虑其 "$S-$ 表现":

$$X_n=X_0+\sum_{k=1}^{n}\gamma_k\Delta S_k, \tag{37}$$

其中 γ_k 是某 $\mathscr{F}_{k-1}-$ 可测随机变量.

现在证明, 由此可见, 存在自筹资金总存量 $\widetilde{\pi}=(\widetilde{\beta},\widetilde{\gamma})$, 使得对于一切 $0\leqslant n\leqslant N$, 特别对于 $f_N=X_N=X_N^{\widetilde{\pi}}$, 有定义 3 所要求的表现:

$$f_N=X_0^{\widetilde{\pi}}+\sum_{k=1}^{n}\widetilde{\gamma}_k\Delta S_k. \tag{38}$$

有了 (37) 式, 现在设 $\widetilde{\gamma}_k=\gamma_k$, 并且定义

$$\widetilde{\beta}_n=X_n-\gamma_nS_n. \tag{39}$$

由 (37) 式可见, 随机变量 $\widetilde{\beta}_n$ 是 $\mathscr{F}_{n-1}-$ 可测的. 这时, 有

$$\begin{aligned}S_{n-1}\Delta\widetilde{\gamma}_n+\Delta\widetilde{\beta}_n&=S_{n-1}\Delta\gamma_n+\Delta X_n-\Delta(\gamma_nS_n)\\&=S_{n-1}\Delta\gamma_n+\gamma_n\Delta S_n-\Delta(\gamma_nS_n)=0.\end{aligned}$$

这样, 根据第三小节, 所建立的总存量 $\widetilde{\pi}=(\widetilde{\beta},\widetilde{\gamma})$ *是自筹的*, 并且 $X_N^{\widetilde{\pi}}=f_N$, 即*无仲裁* $(B,S)-$ *市场是完全的得证*.

于是, 引理 3 得证.□

3) 由引理 3 可见, 为完成定理 2 的证明, 需要证明如下一系列蕴涵关系:

$$\boxed{|\mathbf{M}(\mathbf{P})|=1}\overset{\{3\}}{\Longrightarrow}\boxed{S\text{ 表现}}\overset{\{2\}}{\Longleftrightarrow}\boxed{\text{完全性}}\overset{\{1\}}{\Longrightarrow}\boxed{|\mathbf{M}(\mathbf{P})|=1}.$$

在证明引理 2 的 "必要性" 时, "蕴涵关系 $\{1\}$ " 已经得证; "蕴涵关系 $\{2\}$" 就是引理 2.

为使蕴涵关系 $\{3\}$ 的证明更加清晰, 考虑由 CRR - 模型描绘的 $(B,S)-$ 市场的特殊情形.

上面 (例 1) 曾指出, 对于这样的模型, 鞅测度 $\widetilde{\mathbf{P}}$ 是唯一的 $(|\mathbf{M}(\mathbf{P})|=1)$. 因此, 需要弄清楚, 这里为什么 (关于鞅测度 $\widetilde{\mathbf{P}}$) "*$S-$ 表现*" 成立. 原来在上面已经指出, 关键是 (4) 式中的变量 ρ_n 只有 a 和 b 两个可能值, 而作为这一事实的推论, 条件分布 $\mathbf{P}\{\Delta S\in\cdot|\mathscr{F}_{n-1}\}$ 全部仅集中在两个点上 ("条件两点").

这样, 考虑例 1 中引进的 CRR - 模型, 并且补充假设对于 $1\leqslant n\leqslant N,\mathscr{F}_n=\sigma(\rho_1,\cdots,\rho_n)$, 而 $\mathscr{F}_0=(\varnothing,\Omega)$. 以 $\widetilde{\mathbf{P}}$ 表示由 (27) 式定义在 $(\Omega,\mathscr{F}_N)$ 上的鞅测度.

设 $X = (X_n, \mathscr{F}_n, \widetilde{\mathbf{P}})_{0\leqslant n\leqslant N}$ 是有界鞅, 则存在函数 $g_n = g_n(x_1, \cdots, x_n)$, 使 $X_n(\omega) = g_n(\rho_1(\omega), \cdots, \rho_n(\omega))$, 因此

$$\Delta X_n = g_n(\rho_1, \cdots, \rho_n) - g_{n-1}(\rho_1, \cdots, \rho_{n-1}).$$

由于 $\widetilde{\mathbf{E}}(\Delta X_n|\mathscr{F}_{n-1}) = 0$, 可见

$$\widetilde{p}g_n(\rho_1, \cdots, \rho_{n-1}, b) - \widetilde{q}g_n(\rho_1, \cdots, \rho_{n-1}, a) = g_{n-1}(\rho_1, \cdots, \rho_{n-1}),$$

即

$$\frac{g_n(\rho_1, \cdots, \rho_{n-1}, b) - g_{n-1}(\rho_1, \cdots, \rho_{n-1})}{\widetilde{q}} = \frac{g_{n-1}(\rho_1, \cdots, \rho_{n-1}) - g_n(\rho_1, \cdots, \rho_{n-1}, a)}{\widetilde{p}}. \tag{40}$$

由于 $\widetilde{p} = (r-a)/(b-a), \widetilde{q} = (b-r)/(b-a)$, 故由 (40) 式可见

$$\frac{g_n(\rho_1, \cdots, \rho_{n-1}, b) - g_{n-1}(\rho_1, \cdots, \rho_{n-1})}{b-r} = \frac{g_n(\rho_1, \cdots, \rho_{n-1}, a) - g_{n-1}(\rho_1, \cdots, \rho_{n-1})}{a-r}. \tag{41}$$

设 $\mu_n(\{a\};\omega) = I(\rho_n(\omega) = a), \mu_n(\{b\};\omega) = I(\rho_n(\omega) = b)$; 并且设

$$W_n(\omega, x) = g_n(\rho_1(\omega), \cdots, \rho_{n-1}(\omega), x) - g_{n-1}(\rho_1(\omega), \cdots, \rho_{n-1}(\omega), x),$$
$$W_n^*(\omega, x) = \frac{W_n(\omega, x)}{x-r}.$$

由于这些记号, 可见

$$\Delta X_n(\omega) = W_n(\omega, \rho_n(\omega)) = \int W_n(\omega, x)\mu_n(\mathrm{d}x;\omega) = \int (x-r)W_n^*(\omega, x)\mu_n(\mathrm{d}x;\omega).$$

由于 (41) 式, 可见 $W_n^*(\omega, x)$ 与 x 无关. 如果将 (41) 式的左侧 (或等价地将右侧) 的式子, 记作 $\gamma_n^*(\omega)$, 则得

$$\Delta X_n(\omega) = \gamma_n^*(\omega)(\rho_n(\omega) - r). \tag{42}$$

从而

$$X_n(\omega) = X_0(\omega) + \sum_{k=1}^{n} \gamma_k^*(\omega)[\rho_k(\omega) - r]. \tag{43}$$

易见, 有

$$\Delta\left(\frac{S_n}{B_n}\right) = \frac{S_{n-1}}{B_{n-1}} \cdot \frac{\rho_n - r}{1+r}.$$

由此, 得

$$\rho_n - r = (1+r)\frac{B_{n-1}}{S_{n-1}}\Delta\left(\frac{S_n}{B_n}\right),$$

故由 (43) 式, 可见

$$X_n(\omega) = X_0(\omega) + \sum_{k=1}^{n} \gamma_k(\omega)\Delta\left(\frac{S_k(\omega)}{B_k}\right), \tag{44}$$

其中

$$\gamma_k(\omega) = \gamma_k^*(\omega)(1+r)\frac{B_{k-1}}{S_{k-1}}.$$

序列

$$\frac{S}{B} = \left(\frac{S_n}{B_n}\right)_{0\leqslant n\leqslant N}$$

关于测度 $\widetilde{\mathbf{P}}$ 是鞅. 于是, 上面引进的关系式 (44) 恰好是, X 关于 (基本) $\widetilde{\mathbf{P}}$– 鞅 S/B 的 “S/B– 表现”.

4) 在 CRR – 模型 (其中 $|\mathbf{M}(\mathbf{P})| = 1$) 的蕴涵关系 {3} 的证明中, 关键是变量 ρ_n 只有 a 和 b 两个可能值. 然而, 结果表明, 关于鞅测度 $\widetilde{\mathbf{P}}$ 唯一性的假设是相当强的条件, 以至于由此可以得到变量 $\rho_n = \Delta S_n/S_{n-1}$ 具有 “两点” 的结构: 存在可以预测的量 $a_n = a_n(\omega)$ 和 $b_n = b_n(\omega)$, 使

$$\widetilde{\mathbf{P}}(\rho_n = a_n|\mathscr{F}_{n-1}) + \widetilde{\mathbf{P}}(\rho_n = b_n|\mathscr{F}_{n-1}) = 1. \tag{45}$$

如果相信这条性质, 那么上面所作的 CRR – 模型中 “S/B– 表现” 的证明, 在一般情形下仍然 “适用”. 这样只剩下证明 (45) 式. 在假定读者独立证明 (45) 式的正确性 (练习题 5) 的情况下, 我们仍然引进某些启发性想法, 说明具备鞅测度的唯一性导致 “条件两点” 的结构.

假设 $\mathbf{Q} = \mathbf{Q}(dx)$ 是 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 上的某一概率分布, 而 $\xi = \xi(x)$ 是一坐标给定的随机变量 ($\xi(x) = x$), 并且满足条件: $\mathbf{E_Q}|\xi| < \infty, \mathbf{E_Q}\xi = 0$ (“鞅性”); 而测度 $\mathbf{Q}$ 具有性质: 假如 $\widetilde{\mathbf{Q}}$ 是满足条件: $\mathbf{E}_{\widetilde{\mathbf{Q}}}|\xi| < \infty, \mathbf{E}_{\widetilde{\mathbf{Q}}}\xi = 0$ 的另一测度, 那么必有 $\widetilde{\mathbf{Q}} = \mathbf{Q}$ (“鞅测度的唯一性”).

可以证明, 测度 $\mathbf{Q}$ 的承载子*最多集中在两点* ($a \leqslant 0$ 和 $b \geqslant 0$) 上, 也可能两点都 “粘合” 在 “零” 点 ($a = b = 0$).

5) 上面提到的启发性想法, 使得可以将上面最后表述的、十分近乎情理的命题, 可以叙述如下.

假设测度 $\mathbf{Q}$ 集中在 x_-, x_0, x_+ 等三个点上, 而且三个点是有序的: $x_- \leqslant x_0 \leqslant x_+$, 其质量 (权重) 相应为 q_-, q_0, q_+. 条件 $\mathbf{E_Q}\xi = 0$ 表示

$$q_-x_- + q_0x_0 + q_+x_+ = 0.$$

假如 $x_0 = 0$, 则 $q_-x_- + q_+x_+ = 0$. 设

$$\widetilde{q}_- = \frac{q_-}{2}, \quad \widetilde{q}_0 = \frac{1}{2} + \frac{q_0}{2}, \quad \widetilde{q}_+ = \frac{q_+}{2}, \tag{46}$$

即把在点 x_- 和 x_+ 上的部分质量 q_- 和 q_+ "汲取到" 点 x_0 上.

由 (46) 式可见, 相应的测度 $\widetilde{\mathbf{Q}} \sim \mathbf{Q}$, 且 $\mathbf{E}_{\widetilde{\mathbf{Q}}}\xi = 0$, 而且 $\widetilde{\mathbf{Q}} \neq \mathbf{Q}$.

但是, 这与具有性质 $\mathbf{E}_{\mathbf{Q}}\xi = 0$ 的测度 $\mathbf{Q}$ 的*唯一性*矛盾.

从而, 测度 $\mathbf{Q}$ 不可能集中在 (x_-, x_0, x_+) 三个点上, 且 $x_0 = 0$. 类似地可以基于 "汲取质量" 的思路, 讨论 $x_0 \neq 0$ 的情形. (详见 [100] 的第 V 章 §4e). □

7. 练习题

1. 假设 $\mathbf{P}\{\Delta S_1 = 0\} < 1$, 对于 $N = 1$ 的情形, 证明无仲裁条件等价于不等式 (18) 成立等价.

2. 证明, 第 4 小节引理 1 的证明中, 条件 (19) 排除可能性 2).

3. 证明, 第 5 小节例 1 中的测度 $\widetilde{\mathbf{P}}$ *是鞅测度*, 并且在这种情形下在 $\mathbf{M}(\mathbf{P})$ 类中是唯一的.

4. 讨论第 5 小节例 2 建立的鞅测度的唯一性.

5. 证明, 对于变量 $S_n/B_n (1 \leqslant n \leqslant N)$ 的分布, 由 (B, S)– 模型中的条件 $|\mathbf{M}(\mathbf{P})| = 1$, 得 "条件两点".

§12. 无仲裁模型中与 "套头交易" 有关的核算

1. 购货保留权 (选择权) 进行套头交易 (hedge, 即买即卖), 是有价证卷总存量的动态管理的基本方法之一. 下面通过所谓*选择权合同* (简称*选择权*) 核算的例子, 阐述这种方法某些基本原理和结果.

作为任意有价证券, (金融工程的工具) *选择权* 具有非常高的*风险* . 然而与此同时这些工具 (及其与其他有价证券的组合, 例如, *期货*) 被成功地利用, 不仅为通过 "市场" 价格的变化而获得收入的目的, 而且在 "市场" 价格出现戏剧性变化的情况下是 (选择权的) *保护工具*.

选择权 (option, 选择) 指金融机构发行的有价证券 (合同), 在协议的时间内或者在事先约定条件下的时间, 赋予买方购买或出售一定的价值 (例如, 股票, 债券, 货币) 的*权限*.

设计选择权核算的基本问题之一, 在于按何种价格出售选择权? 显然, 卖方希望获得 "尽量多", 而买方希望付出 "尽量少". 问什么价格是 "正确的", "合理的", 是买方和卖方应该都可以接受的价格?

自然, 这一 "合理的" 价格应该是 "明智的". 具体地说, 买方应该明白, 用较低的价格购买选择权, 不能得到卖方的兑现自己义务的承诺, 因为它所得到的 "奖励", 有可能简直不足以保障构成 "供货委托" 总存量.

与此同时, 这些 "奖励" 不应给卖方提供 "free lunch (免费午餐)" 一类的仲裁机会, 即得到*无风险收入*的可能性.

在给出应如何理解选择权的“正确”价格的定义之前, 我们首先考虑选择权一些被普遍采用的分类.

2. 购货保留权 (选择权) 的类型 设 $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{0\leqslant n\leqslant N}, \mathbf{P})$ 是过滤概率空间, 而 $\mathscr{F}_0 = \{\varnothing, \Omega\}$, 且 $\mathscr{F}_N = \mathscr{F}$. 考虑定义在给定概率空间上, 在时间 $n = 0, 1, \cdots, N$ 运转的 (B, S)– 市场, 其中 $B = (B_n)_{0\leqslant n\leqslant N}, S = (S_n)_{0\leqslant n\leqslant N}$.

下面考虑的选择权, 建立在股票上的情形, 假设其价格由序列 $S = (S_n)_{0\leqslant n\leqslant N}$ 描绘的.

选择权按的执行时间分为两种类型: *欧洲型*和*美国型*.

如果选择权可以表现为仅在合同固定的时间 N 进行, 则称 N 为执行时间, 这样的选择权称为*欧洲型的*.

如果选择权可以表现为, 可以在任意马尔可夫时间或者停止时间 $\tau = \tau(\omega)$ 进行 (见 §1 定义 3), 且 $\tau = \tau(\omega)$ 在合同的条件中, 预先说明的集合 $\{0, 1, \cdots, N\}$ 中的值, 则所考虑的选择权是*美国型的*.

按照通用的术语, 区分如下两类选择权:

(1) *买方选择权* (call option①, 由此产生了的术语 опцион–колл (购买权))

(2) *卖方选择权* (put option②, 由此产生了的术语 опцион–пут (出售权)).

为确定计, 我们考虑*欧洲型*标准选择权的例子.

这样的选择权由两份合同体现: 执行时间 N 和购买价格 K (对于买方), 或出售价格 K (对于卖方).

假如, 万一在时间 N “市场” 价格 $S_N > K$, 则根据合同的条件, 选择权的买方有权按价格 K 购买股票. 买方同样有权按 “市场” 价格 S_N 出售, 这样可获利 $S_N - K$. 假如结果为 $S_N < K$, 则协议商定的按价格 K 的购买权应毫无疑义地采用, 因为可以用较低于 “市场” 价的价格 S_N 购入股票.

于是, 综合这两种情形可见, 买方在时间 N 的收入决定于量:

$$f_N = (S_N - K)^+, \tag{1}$$

其中对于任意实数 $a, a^+ = \max(a, 0)$. 其 “纯” 收入等于 f_N 减去他支付给股票卖方的 “奖励”.

类似地, 出售权卖方的收入决定于量:

$$f_N = (K - S_N)^+, \tag{2}$$

3. 完全市场和不完全市场 在定义无仲裁 (B, S)– 市场的 “合理” 价值, 应当区分两种情形: *完全市场和不完全市场*.

①买证券的特权, 买期货的权利. 俄语译为 “опцион–колл ”. —— 译者

②买期货的权利. 俄语译为 “опцион–пут”. —— 译者

定义 1　设 (B,S)– 市场是无仲裁的和完全的. 完全套头交易的价格, 即量

$$\mathbb{C}(f_N;\mathbf{P}) = \inf\{x : \exists\pi, \text{ 使 } X_0^\pi = x, \quad X_N^\pi = f_N(\mathbf{P}-\text{a.c.})\}, \tag{3}$$

称做欧洲型选择权的 "合理" 价格, 其中 f_N 是 $\mathscr{F}_N$– 可测有界 (非负) 支付函数.

关于这一定义我们指出, 如果依 $\mathbf{P}$– 概率 1 有 $X_N^\pi \geqslant f_N$, 则称总存量 π 为支付委托的费用. 由 §11 可见, 对于完全无仲裁市场的情形, 存在有界 "支付费用" 的完全套头交易 π, 即满足 $X_N^\pi = f_N(\mathbf{P}-\text{a.c.})$ 的 π. 这恰好说明, 为什么在定义 3 中, 要考虑具有性质 $X_N^n = f_N$ $(\mathbf{P}-\text{a.c.})$ 的 (非空) 总存量类.

对于不完全无仲裁市场的情形, 自然地有如下定义.

定义 2　设 (B,S)– 市场是无仲裁的. 超级套头交易的价格, 即价格

$$\mathbb{C}(f_N;\mathbf{P}) = \inf\{x : \exists\pi, \quad \text{使 } X_0^\pi = x, \quad X_N^\pi \geqslant f_N(\mathbf{P}-\text{a.c.})\}, \tag{4}$$

称做具有 $\mathscr{F}_N$– 可测有界 (非负) 支付函数 f_N 的, 欧洲型选择权的 "合理" 价格.

注意, 该定义是适定的: 对于任何有界函数 f_N, 一定存在具有初始资本 x 的总存量 π, 使 $X_N^\pi \geqslant f_N(\mathbf{P}-\text{a.c.})$.

4. 价格公式　现在引进 $\mathbb{C}(f_N;\mathbf{P})$ 的公式, 并且对于完全市场的情形给予证明, 而对于不完全市场的情形, 介绍专门的文献 ([100] 第 VI 章 §1).

定理 1　1) 对于完全的无仲裁 (B,S)– 市场的情形, 支付函数为 f_N 的欧洲型选择权的 "合理" 价格, 决定于公式

$$\mathbb{C}(f_N;\mathbf{P}) = B_0\mathbf{E}_{\widetilde{\mathbf{P}}}\frac{f_N}{B_N}, \tag{5}$$

其中 $\mathbf{E}_{\widetilde{\mathbf{P}}}$ 是对 (唯一) 鞅测度 $\widetilde{\mathbf{P}}$ 的数学期望.

2) 对于一般不完全的无仲裁 (B,S)– 市场的情形, 支付函数为 f_N 的欧洲型选择权的 "合理" 价格, 决定于公式

$$\mathbb{C}(f_N;\mathbf{P}) = \sup_{\widetilde{\mathbf{P}}\in\mathbf{M}(\mathbf{P})} B_0\mathbf{E}_{\widetilde{\mathbf{P}}}\frac{f_N}{B_N}, \tag{6}$$

其中 sup 在一切鞅测度 $\mathbf{M}(\mathbf{P})$ 的集合上来求.

证明　1) 设 π 是某完全套头交易, 其中 $X_0^\pi = x, X_N^\pi = f_N(\mathbf{P}-\text{a.c.})$. 那么, 有 (见 §11 的 (15) 式)

$$\frac{f_N}{B_N} = \frac{X_N^\pi}{B_N} = \frac{x}{B_0} + \sum_{k=1}^N \gamma_k\Delta\left(\frac{S_k}{B_k}\right), \tag{7}$$

因而, 由于 §1 的定理 3, 得

$$\mathbf{E}_{\widetilde{\mathbf{P}}}\frac{f_N}{B_N} = \frac{x}{B_0}, \tag{8}$$

因为, 由于鞅变换

$$\left(\frac{x}{B_0} + \sum_{k=1}^n \gamma_k\Delta\left(\frac{S_k}{B_k}\right)\right)_{1\leqslant n\leqslant N},$$

可见在“最末”时间 N, 有

$$\frac{x}{B_0}+\sum_{k=1}^{N}\gamma_k\Delta\left(\frac{S_k}{B_k}\right)=\frac{f_N}{B_N}\geqslant 0. \tag{9}$$

注意,(8) 式的左侧不依赖于, 所考虑的初始值为 $X_0^\pi=x$ 的套头交易 π. 假如现在 π' 是初始值为 $X_0^{\pi'}$ 的另外一起套头交易, 则根据 (8) 式其值仍等于 $B_0\mathbf{E}_{\widetilde{\mathbf{P}}}(f_N/B_N)$. 由此显然, 对于一切完全套头交易, 初始值 x 是同一个. 于是, 就证明了 (5) 式.

2) 我们在这里只证明不等式

$$\sup_{\widetilde{\mathbf{P}}\in\mathbf{M}(\mathbf{P})} B_0\mathbf{E}_{\widetilde{\mathbf{P}}}\frac{f_N}{B_N}\leqslant\mathbb{C}(f_N;\mathbf{P}). \tag{10}$$

(相反不等式的证明, 要用到所谓“选择权”的分解, 而这已经超出本书的范围; 见 [100] 第 VI 章的 §1c) 和 §2d)).

假设对套头交易 π , 有 $X_0^\pi=x, X_N^\pi\geqslant f_N(\mathbf{P}-\text{a.c.})$.

那么, 由 (7) 式可见

$$\frac{x}{B_0}+\sum_{k=1}^{N}\gamma_k\Delta\left(\frac{S_k}{B_k}\right)\geqslant\frac{f_N}{B_N}\geqslant 0.$$

因此, 对于任何 测度 $\widetilde{\mathbf{P}}\in\mathbf{M}(\mathbf{P})$, 有

$$B_0\mathbf{E}_{\widetilde{\mathbf{P}}}\frac{f_N}{B_N}\leqslant x$$

(对照 (8) 式和 (9) 式). 于是, 如果该式的左侧对于所有 $\widetilde{\mathbf{P}}\in\mathbf{M}(\mathbf{P})$ 求 sup, 则得所要证明的 (10) 式. □

5. 美国型购货保留权 (选择权) 考虑涉及美国型选择权的一些定义和结果. 对于这样的选择权需要作如下假设: 给定涉及时间 N 的不只一个支付函数 f_N, 而是整整一组函数 $f_0, f_1,\cdots,f_N$; 这些支付函数的含义是, 如果选择权表明买方出现在时间 n, 则 (卖方给买方选择权的) 相应的支付决定于 ($\mathscr{F}_n$− 可测) 函数 $f_n=f_n(\omega)$.

以 $\tau=\tau(\omega)$ 表示值域为 $\{0,1,\cdots,N\}$ 的马尔可夫时间. 假如选择权的买方决定在时间 $\tau=\tau(\omega)$ 提出要求执行选择权, 则支付函数等于 $f_{\tau(\omega)}(\omega)$, 因而有价证券 π 选择权的卖方永远应当预见到, 使得对于任何 τ, 都能满足套头交易的如下条件: $X_\tau^\pi\geqslant f_\tau(\mathbf{P}-\text{a.c.})$.

这说明了如下定义的合理性.

定义 3 设 (B,S)− 市场是无仲裁的, 而 $f=(f_n)_{0\leqslant n\leqslant N}$ 是 $\mathscr{F}_n$− 可测非负支付函数 f_n 系. 超级套头交易的最高价格,即价格

$$\overline{\mathbb{C}}(f;\mathbf{P})=\inf\{x:\exists\pi, X_0^\pi=x, X_n^\pi\geqslant f_n(\mathbf{P}-\text{a.c.}), 0\leqslant n\leqslant N\}, \tag{11}$$

称做以 $f=(f_n)_{0\leqslant n\leqslant N}$ 为支付函数系的、美国型选择权的“合理”价格.

对于美国型选择权的情形, 我们 (不加证明地) 引进定理 1 的如下类似.

定理 2 1) 对于完全的无仲裁 $(B,S)-$ 市场的情形, 以 $f=(f_n)_{0\leqslant n\leqslant N}$ 为支付函数系的、美国型选择权的 "合理" 价格, 决定于公式

$$\overline{\mathbb{C}}(f;\mathbf{P})=\sup_{\tau\in\mathfrak{M}_0^N} B_0\mathbf{E}_{\widetilde{\mathbf{P}}}\frac{f_\tau}{B_\tau}, \tag{12}$$

其中 $\mathfrak{M}_0^N=\{\tau:\tau\leqslant N\}$ 是 (关于 $(\mathscr{F}_n)_{0\leqslant n\leqslant N}$) 的停时类, 而 $\widetilde{\mathbf{P}}$ 是唯一鞅测度. $\mathbf{E}_{\widetilde{\mathbf{P}}}$ 是对 (唯一) 鞅测度 $\widetilde{\mathbf{P}}$ 的数学期望.

2) 对于一般不完全的无仲裁 $(B,S)-$ 市场的情形, 支付函数为 $f=(f_n)_{0\leqslant n\leqslant N}$ 的美国型选择权的 "合理" 价格, 决定于公式

$$\overline{\mathbb{C}}(f;\mathbf{P})=\sup_{\tau\in\mathfrak{M}_0^N,\widetilde{\mathbf{P}}\in\mathbf{M}(\mathbf{P})} B_0\mathbf{E}_{\widetilde{\mathbf{P}}}\frac{f_\tau}{B_\tau}, \tag{13}$$

其中 $\mathbf{M}(\mathbf{P})$ 是一切鞅测度 $\widetilde{\mathbf{P}}$ 的集合.

证明见 [100] 第 Ⅵ 章的 §2c.

6. 欧洲型购货保留权 (选择权) 上面的两个定理回答选择权如何决定其 "合理" 价格的问题.

在得到 "奖励" $\mathbb{C}(f_N;\mathbf{P})$ 和 $\overline{\mathbb{C}}(f;\mathbf{P})$ 之后, 选择权的卖方如何建立套头存量 π^* 的问题也同样重要.

为简便计, 我们的叙述仅限于考虑欧洲型购货选择权完全 $(B,S)-$ 市场的情形.

定理 3 设 $(B,S)-$ 市场是无仲裁的和完全的.

存在自筹总存量 $\pi^*(\beta^*,\gamma^*)$, 其初始资本为 $X_0^{\pi^*}=\mathbb{C}(f_N;\mathbf{P})$, 并且实行完全支付义务 f_N 的套头交易:

$$X_0^{\pi^*}=f_N(\mathbf{P}-\text{a.c.}).$$

资本 $X_n^{\pi^*}=\beta_n^*B_n+\gamma_n^*S_n(0\leqslant n\leqslant N)$ 的变动决定于如下公式

$$X_n^{\pi^*}=B_n\mathbf{E}_{\widetilde{\mathbf{P}}}\left(\frac{f_N}{B_N}\middle|\mathscr{F}_n\right). \tag{14}$$

自筹总存量 $\pi^*=(\beta^*,\gamma^*)$ 中的分量 $\gamma^*=(\gamma_n^*)_{0\leqslant n\leqslant N}$, 根据 $X^{\pi^*}=(X_n^{\pi^*})_{0\leqslant n\leqslant N}$ 的值由公式

$$\Delta\left(\frac{X_n^{\pi^*}}{B_n}\right)=\gamma_n^*\Delta\left(\frac{S_n}{B_n}\right), \tag{15}$$

来求, 而分量 $\beta^*=(\beta_n^*)_{0\leqslant n\leqslant N}$ 的值由公式

$$X_n^{\pi^*}=\beta_n^*B_n+\gamma_n^*S_n, \tag{16}$$

来求.

定理的证明可以直接仿照 §11 引理 2 中蕴涵关系

$$\text{“完全性”} \Rightarrow \text{“}\frac{S}{B} - \text{表现”}$$

的证明得到. 为此, 只需将上面蕴涵关系的证明用于鞅

$$m = (m_n)_{0 \leqslant n \leqslant N}, \quad m_n = \mathbf{E}_{\widetilde{\mathbf{P}}}\left(\frac{f_N}{B_N}\middle|\mathscr{F}_n\right).$$

7. 选择权实际核算的例 作为选择权实际核算的例子, 考虑由 CCR – 模型

$$\begin{aligned} B_n &= B_{n-1}(1+r), \\ S_n &= S_{n-1}(1+\rho_n), \end{aligned} \tag{17}$$

描绘的 (B,S)– 市场, 其中 $\rho_1, \cdots, \rho_N$ 是只有 a 和 $b(-1 < a < r < b)$ 两个可能值的、独立同分布随机变量.

这样的市场是无仲裁的和完全的 (见 §11 练习题 3), 且与其相联系的鞅测度 $\widetilde{\mathbf{P}}$ 满足: $\widetilde{\mathbf{P}}\{\rho_n = b\} = \widetilde{p}, \widetilde{\mathbf{P}}\{\rho_n = a\} = \widetilde{q}$, 其中

$$\widetilde{p} = \frac{r-a}{b-a}, \quad \widetilde{q} = \frac{b-r}{b-a}, \tag{18}$$

(见 §11 第 5 小节例 1).

根据定理 1 的 (5) 式, 对于所考虑的 (B,S)– 市场, "合理" 价格为

$$\mathbb{C}(f_N;\mathbf{P}) = \mathbf{E}_{\widetilde{\mathbf{P}}}\frac{f_N}{(1+r)^N}. \tag{19}$$

故根据定理 3 求完全套头交易总存量 $\pi^*(\beta^*, \gamma^*)$, 需要首先计算

$$X_0^{\pi^*} = \mathbf{E}_{\widetilde{\mathbf{P}}}\left(\frac{f_N}{(1+r)^N}\middle|\mathscr{F}_n\right), \tag{20}$$

其中 $\mathscr{F}_n = \sigma(\rho_1, \cdots, \rho_n), 1 \leqslant n \leqslant N$, 且 $\mathscr{F}_0 = \{\varnothing, \Omega\}$; 然后由公式 (15) 和 (16) 求 β^* 和 γ^*.

由于 $X_0^{\pi^*} = \mathbb{C}(f_N;\mathbf{P})$, 可见这一切归结为, 对于 $n = 0, 1, \cdots, N$, 求 (20) 式右侧的条件数学期望.

假设 $\mathscr{F}_n$– 可测函数 f_N 有 "马尔可夫" 结构, 即 $f_N = f(S_N)$, 其中 $f = f(x)$ 是 $x \geqslant 0$ 的某一非负函数.

记

$$F_n(x;p) = \sum_{k=0}^{n} f(x(1+b)^k(1+a)^{n-k})\mathrm{C}_n^k p^k(1-p)^{n-k}. \tag{21}$$

由于

$$\prod_{n<k\leqslant N}(1+\rho_k) = (1+b)^{\Delta_N - \Delta_n}(1+a)^{(N-n)-(\Delta_N-\Delta_n)},$$

其中 $\Delta_k = \delta_1 + \cdots + \delta_n$, 而 $\delta_k = (p_k - a)/(b-a)$, 可见

$$\mathbf{E}_{\widetilde{\mathbf{P}}} f\left(x \prod_{n<k\leqslant N}(1+\rho_k)\right) = F_{N-n}(x,\widetilde{p}), \quad \text{其中 } \widetilde{p} = \frac{r-a}{b-a}. \tag{22}$$

而考虑到

$$S_N = S_n \prod_{n<k\leqslant N}(1+\rho_k),$$

故由式 (21) 式和 (22) 式, 得

$$X_0^{\pi^*} = \mathbf{E}_{\widetilde{\mathbf{P}}}\left(\frac{f_N}{(1+r)^N}\middle|\mathscr{F}_n\right) = (1+r)^{-N}F_{N-n}(S_n,\widetilde{p}). \tag{23}$$

特别,

$$\mathbb{C}(f_N;\mathbf{P}) = X_0^{\pi^*} = (1+r)^{-N}F_N(S_0,\widetilde{p}). \tag{24}$$

于是, 由 (15) 式, 并注意到 (23) 式, 可见

$$\gamma_n^* = \Delta\left(\frac{X_n^{\pi^*}}{B_n}\right)\Big/\Delta\left(\frac{S_n}{B_n}\right)$$

决定于

$$\gamma_n^* = (1+r)^{-(N-n)}\frac{F_{N-n}(S_{n-1}(1+b);\widetilde{p}) - F_{N-n}(S_{n-1}(1+a);\widetilde{p})}{S_{n-1}(b-a)}. \tag{25}$$

为求 β_n^*, 注意到 $B_{n-1}\Delta\beta_n^* + S_{n-1}\Delta\gamma_n^* = 0$. 所以

$$X_{n-1}^{\pi^*} = \beta_n^* B_{n-1} + \gamma_n^* S_{n-1}, \tag{26}$$

因此

$$\beta_n^* = \frac{X_{n-1}^{\pi^*} - \gamma_n^* S_{n-1}}{B_{n-1}}. \tag{27}$$

由 (23) 式和 (25) 式, 可见

$$\begin{aligned}\beta_n^* = \frac{1}{B_N}\bigg\{&F_{N-n+1}(S_{n-1};\widetilde{p}) - \frac{1+r}{1+b}[F_{N-n}(S_{n-1}(1+b);\widetilde{p})\\ &-F_{N-n}(S_{n-1}(1+a);\widetilde{p})]\bigg\}.\end{aligned} \tag{28}$$

最后, 我们讨论一下, 例如, 在买方的标准选择权 (购买权) 的情形下, 即当函数 $f_N = (S_N - K)^+$ 时, "合理" 价格 $\mathbb{C}(f_N,\mathbf{P})$ 的公式具有何种形式.

设 $K_0 = K_0(a,b,N;s_0/K)$ 是满足如下条件的最小整数:

$$S_0(1+a)^N\left(\frac{1+b}{1+a}\right)^{K_0} > K, \tag{29}$$

即设

$$K_0 = 1 + \left[\ln\left(\frac{K}{S_0(1+a)^N}\right) \Big/ \ln\left(\frac{1+b}{1+a}\right) \right], \tag{30}$$

其中 $[x]$ 表示 x 的整数部分.

如果记

$$p^* = \frac{1+b}{1+r}\widetilde{P}, \quad 其中 \quad \widetilde{P} = \frac{r-a}{b-a},$$

而

$$\mathbb{B}(K_0, N, p) = \sum_{k=K_0}^{N} \mathrm{C}_N^k p^k (1-p)^{N-k}, \tag{31}$$

则对于标准购买权, 由 (24) 式不难推导出“合理”价格 (记作 $\mathbb{C}_N$) 的如下 (考科斯 – 罗斯 – 鲁宾斯坦) 公式:

$$\mathbb{C}_N = S_0\mathbb{B}(K_0, N; p^*) - K(1+r)^{-N}\mathbb{B}(K_0, N; \widetilde{p}). \tag{32}$$

当 $K_0 > N$ 时 $\mathbb{C}_N = 0$.

注 由于

$$(K - S_N)^+ = (S_N - K)^+ - S_N + K,$$

可见卖方的标准选择权 (出售权) “合理”价格 $\mathbb{P}_N (= \mathbb{C}(f_N, \mathbf{P})$, 其中 $f_N = (K - S_N)^+)$ 决定于如下公式:

$$\mathbb{P}_N = \widetilde{\mathbf{E}}(1+r)^{-N}(K - S_N)^+ = \mathbb{C}_N - \widetilde{\mathbf{E}}(1+r)^{-N}S_N + K(1+r)^{-N}.$$

因为 $\widetilde{\mathbf{E}}(1+r)^{-N}S_N = S_0$, 所以显然有“平价购买 – 出售权”恒等式:

$$\mathbb{P}_N = \mathbb{C}_N - S_0 + K(1+r)^{-N}. \tag{33}$$

8. 练习题

1. 对于 §11 第 5 小节例 2 中的 (B,S)– 市场模型, 求标准选择权的出售权的“合理”价格: $\mathbb{C}(f_N, \mathbf{P})$, 其中 $f_N = (S_N - K)^+$.
2. 试证明 (10) 式中相反不等式的正确性.
3. 证明 (12) 式,并试证明 (13) 式.
4. 证明 (23) 式详细类似的结论.
5. 证明 (25) 式和 (28) 式.
6. 对 (32) 式进行详细的推导.

§13. 最优停时问题. 鞅方法

1. "合理" 价格 在这一节中, 我们假设给定一固定的概率空间 $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 0}, \mathbf{P})$ 和 $\mathscr{F}_0 = (\varnothing, \Omega)$. 设 $\tau = \tau(\omega)$ 是马尔可夫时间 (或停止时间), 其值域是集合 $\{0, 1, \cdots, N\}$. 设 $\mathfrak{M}_0^N$ 是由属于集合

$$\{\omega : \tau(\omega) = n\} \in \mathscr{F}_n \tag{1}$$

的值域为 $\{0, 1, \cdots, N\}$ 的随机变量 $\tau = \tau(\omega)$ 类. 在描绘美国型选择权 "合理" 价格时, 我们已经遇到过有关停止时间的例子. 具体地说, §12 的 (12) 式表明, 对于简化了的条件: $B_n = 1, 0 \leqslant n \leqslant N$ 和 $\widetilde{\mathbf{P}} = \mathbf{P}$, 为求这一价格, 需要寻找量 (也称做 "价格")

$$V_0^N = \sup_{\tau \in \mathfrak{M}_0^N} \mathbf{E} f_\tau, \tag{2}$$

其中 $f = (f_0, f_1 \cdots, f_N)$ 是 $\mathscr{F}_n$– 可测非负函数 f_n 的序列.

设 $\tau = \tau(\omega) \in \mathfrak{M}_0^N$ 是马尔可夫时间. 与问题 (1) 同时, 对寻求量 ("价格")

$$V_0^\infty = \sup_{\tau \in \mathfrak{M}_0^\infty} \mathbf{E} f_\tau \tag{3}$$

感兴趣, 其中 $\mathfrak{M}_0^N = \{\tau : \tau < \infty\}$, 而 $f = (f_0, f_1, \cdots)$ 是 $\mathscr{F}_n$– 可测随机变量 f_n, $n \geqslant 0$ 的随机序列, 且 $\mathbf{E}|f_\tau| < \infty$.

像 (2)式的情形一样, 对于情形 (3), 除寻求 "价格" V_0^N 和 V_0^∞ 之外, (若上确界存在) 还要寻求达到上确界的时间.

在许多问题中, 允许考虑可以取 $+\infty$ 为值的马尔可夫时间也是适宜的. 在这种情形下, 讨论 $\mathbf{E} f_\tau$ 时应当约定如何理解 f_∞. 一种自然的方法是把 f_∞ 视为的 $\overline{\lim}_n f_n$ 的值. 另一种方法是, 允许对于 τ 值和无限大, 把 "价格" 定义为

$$\overline{V}_0^\infty = \sup_{\tau \in \overline{\mathfrak{M}}_0^\infty} \mathbf{E} f_\tau I(\tau < \infty), \tag{4}$$

其中 $\overline{\mathfrak{M}}_0^\infty$ 是马尔可夫时间类, 即一切马尔可夫时间的集合 $\overline{\mathfrak{M}}_0^\infty = \{\tau : \tau \leqslant \infty\}$. 显然, 若 $f_\infty = 0$, 则

$$\overline{V}_0^\infty = \sup_{\tau \in \overline{\mathfrak{M}}_0^\infty} \mathbf{E} f_\tau$$

(对照 §1 的第 3 小节).

我们在下面将只考虑问题 (2). (关于 $N = \infty$ 的情形, 见第 VIII 章 §9). 假如不考虑序列 $f = (f_0, f_1, \cdots, f_N)$ 的具体购造, 则问题 (2) 和 (3) 的解法是下面描绘的 "鞅" 方法. (不失普遍性, 我们以后都假设对于一切 $n \leqslant N$, 有 $\mathbf{E}|f_n| < \infty$, 并且每次不再特别说明).

2. 价格最高的时间 设 $N < \infty$. 现在考虑的情形可以看成 "向后归纳法", 这里用下面的方式实现.

与价格 V_0^N 同时引进 "价格"

$$V_n^N = \sup_{\tau \in \mathfrak{M}_n^N} \mathbf{E} f_\tau, \tag{5}$$

其中 $\mathfrak{M}_n^N = \{\tau : n \leqslant \tau \leqslant N\}$ 是对于一切 $\omega \in \Omega$, 满足 $n \leqslant \tau(\omega) \leqslant N$ 的停时类.

对于 $n = N-1, \cdots, 0$, 同样直观地引进随机序列 $v^N = (v_n^N)_{0 \leqslant n \leqslant N}$, 其中

$$v_N^N = f_N, \quad v_n^N = \max\{f_n, \mathbf{E}(v_{n+1}^N | \mathscr{F}_n)\}. \tag{6}$$

对于 $0 \leqslant n \leqslant N$, 设

$$\tau_n^N = \min\{n \leqslant k \leqslant N : f_k = v_k^N\}. \tag{7}$$

下面的命题可以完整地描绘, 利用所引进的量, 求解 (2) 式 (5) 式的最优停止的问题.

定理 1 假设对于序列 $f = (f_0, f_1, \cdots, f_N)$, 随机变量 $f_n, n \geqslant 0$ 为 $\mathscr{F}_n-$ 可测.

1) 对于每一个 $n (0 \leqslant n \leqslant N)$, 时间

$$\tau_n^N = \min\{n \leqslant k \leqslant N : v_k^N = f_k\}. \tag{8}$$

在类 $\mathfrak{M}_n^N$ 中是最优的:

$$\mathbf{E} f_{\tau_n^N} = \sup_{\tau \in \mathfrak{M}_n^N} \mathbf{E} f_\tau \quad (= V_n^N). \tag{9}$$

2) 时间 $\tau_n^N (0 \leqslant n \leqslant N)$ 在如下的 "条件" 意义上也是最优的: $(\mathbf{P} - \text{a.c.})$

$$\mathbf{E}(f_{\tau_n^N} | \mathscr{F}_n) = \operatorname*{ess\,sup}_{\tau \in \mathfrak{M}_n^N} \mathbf{E}(f_\tau | \mathscr{F}_n). \tag{10}$$

"随机价格" $\operatorname*{ess\,sup}_{\tau \in \mathfrak{M}_n^N} \mathbf{E}(f_\tau | \mathscr{F}_n)$ 等于 v_n^N:

$$\operatorname*{ess\,sup}_{\tau \in \mathfrak{M}_n^N} \mathbf{E}(f_\tau | \mathscr{F}_n) = v_n^N \quad (\mathbf{P} - \text{a.c.}) \tag{11}$$

且

$$V_n^N = \mathbf{E} v_n^N. \tag{12}$$

如果 $n = 0$, 则

$$V_0^N = v_0^N. \tag{13}$$

如果 $n = N$, 则

$$V_N^N = \mathbf{E} f_N. \tag{14}$$

3. 随机变量族的本质上确界　在开始证明之前我们回忆, 关于 (10) 式中使用的 $\mathscr{F}_n$– 可测随机变量族 $\{\xi_\alpha(\omega), \alpha \in \mathfrak{U}\}$ 的、本质上确界 $\operatorname{ess\,sup}_{\alpha\in\mathfrak{U}} \xi_\alpha(\omega)$ 的概念.

下面的情况决定了引进这一概念的必要性: 在不可数集合 $\mathfrak{U}$ 的情形下, 函数 $\sup_{\alpha\in\mathfrak{U}} \xi_\alpha(\omega)(\omega \in \Omega)$ 一般有可能是 $\mathscr{F}$– 不可测的.

事实上, 对于任意 $c \in \mathbb{R}$,

$$\left\{\omega : \sup_{\alpha\in\mathfrak{U}} \xi_\alpha(\omega) \leqslant c\right\} = \bigcap_{\alpha\in\mathfrak{U}} \{\omega : \xi_\alpha(\omega) \leqslant c\}.$$

这里, 集合 $A_\alpha = \{\omega : \xi_\alpha(\omega) \leqslant c\} \in \mathscr{F}$ (即 A_α 是事件). 然而, 由于集合 $\mathfrak{U}$ 的不可数性, 故不能保障 $\bigcap_{\alpha\in\mathfrak{U}} A_\alpha = \{\omega : \xi_\alpha(\omega) \leqslant c\} \in \mathscr{F}$.

定义　设 $\{\xi_\alpha(\omega) : \alpha \in \mathfrak{U}\}$ 是随机变量族 (即值域为 $(-\infty, +\infty)$ 的 $\mathscr{F}_n$– 可测函数族). 称广义随机变量 $\xi(\omega)$ (即值域为 $(-\infty, +\infty)$ 的 $\mathscr{F}_n$– 可测函数) 是随机变量族 $\{\xi_\alpha(\omega) : \alpha \in \mathfrak{U}\}$ 的本质上确界, 记作 $\xi(\omega) = \operatorname{ess\,sup}_{\alpha\in\mathfrak{U}} \xi_\alpha(\omega)$, 如果

a) 对于一切 $\alpha \in \mathfrak{U}, \xi(\omega) \geqslant \xi_\alpha(\omega)(\mathbf{P} - \text{a.c.})$,

b) 由于对于, (广义) 随机变量 $\eta(\omega)$, 对一切 $\alpha \in \mathfrak{U}$ 满足 $\eta(\omega) \geqslant \xi_\alpha(\omega)(\mathbf{P} - \text{a.c.})$, 可见 $\xi(\omega) \leqslant \eta(\omega)(\mathbf{P} - \text{a.c.})$.

换句话说, 在一切控制随机变量 $\xi_\alpha(\omega), \alpha \in \mathfrak{U}$ 的 (广义) 随机变量中, $\xi(\omega)$ 是最小的 (广义) 随机变量.

当然, 首先应该证明这一定义富有内容, 而这由下面的引理可以看到.

引理　*对于任意随机变量族 $\{\xi_\alpha(\omega) : \alpha \in \mathfrak{U}\}$, 存在具有定义中性质 a) 和 b) 的随机变量 $\xi(\omega)$ (记作 $\xi(\omega) = \operatorname{ess\,sup}_{\alpha\in\mathfrak{U}} \xi_\alpha(\omega)$, 一般为广义随机变量).*

存在具有同样性质的子集 $\mathfrak{U}_0 \subseteq \mathfrak{U}$, 并且作为这样的随机变量, 可以取随机变量

$$\xi(\omega) = \sup_{\alpha\in\mathfrak{U}_0} \xi_\alpha(\omega).$$

证明　首先假设所有变量 $\xi_\alpha(\omega), \alpha \in \mathfrak{U}$,一致有界 $(|\xi_\alpha(\omega)| \leqslant c, \omega \in \Omega, \alpha \in \mathfrak{U})$.

设 A 是下标 $\alpha \in \mathfrak{U}$ 的有限集合. 记

$$S(A) = \left(\max_{\alpha\in A} \xi_\alpha(\omega)\right).$$

其次, 设 $S = \sup S(A)$, 其中上确界对一切有限子集 $A \subseteq \mathfrak{U}$ 来求.

对于 $n \geqslant 1$, 以 A_n 表示满足

$$\mathbf{E}\left(\max_{\alpha\in A} \xi_\alpha(\omega)\right) \geqslant S - \frac{1}{n}$$

的有限集合 A_n.

记 $\mathfrak{U}_0 = \bigcap_{n\geqslant 1} A_n$. 由于这一集合的可数, 函数

$$\xi(\omega) = \sup_{\alpha\in\mathfrak{U}_0} \xi_\alpha(\omega)$$

$\mathscr{F}_n-$ 可测, 因此是随机变量. (注意, 由于 $|\xi(\omega)| \leqslant c$, 可见 $\xi(\omega)$ 是普通的, 而不是广义随机变量).

由随机变量 $\xi(\omega)$ 如上的构造可见 (练习题 1), 它满足上面定义中的条件 a) 和 b).

因而, 在随机变量族 $\{\xi_\alpha(\omega) : \alpha \in \mathfrak{U}\}$ 一致有界的情形下, 就证明了上确界的存在性.

在一般情形下, 需要首先由随机变量 $\xi_\alpha(\omega)$, 转换为有界随机变量 $\widetilde{\xi}_\alpha(\omega) = \arctan\xi_\alpha(\omega)$, 使之满足 $|\widetilde{\xi}_\alpha(\omega)| \leqslant \pi/2, \alpha\in\mathfrak{U}, \omega\in\Omega$, 然后建立 $\widetilde{\xi}(\omega) = \operatorname{ess}\sup\limits_{\alpha\in\mathfrak{U}}\widetilde{\xi}_\alpha(\omega)$.

随机变量 $\xi_\alpha(\omega) = \tan\widetilde{\xi}_\alpha(\omega)$ 满足本质上确界定义 (练习题 1) 的条件 a) 和 b).

4. 定理 1 的证明

证明 固定上标 N, 而为简便计现在将其略去.

如果 $n = N$, 则 $v_N = f_N$ 且 $\tau_N = N$, 故性质 (9) ~ (12), (14) 显然. 现在用归纳法证明.

设对于 $n = N, N-1, \cdots, k$, 定理的论断已经得证. 现在证明对于 $n = k-1$, 定理的论断也成立.

设 $\tau\in\mathfrak{M}_{k-1}(=\mathfrak{M}_{k-1}^N)$ 和 $A\in\mathscr{F}_{k-1}$. 定义时间 $\overline{\tau}\in\mathfrak{M}_k$, 设 $\overline{\tau} = \max(\tau, k)$. 由于 $\overline{\tau}\in\mathfrak{M}_k$ 且事件 $\{\tau\geqslant k\}\in\mathscr{F}_{k-1}$, 可见

$$\begin{aligned}\mathbf{E}[I_A f_\tau] &= \mathbf{E}[I_{A\cap\{\tau=k-1\}}f_\tau] + \mathbf{E}[I_{A\cap\{\tau\geqslant k\}}f_\tau] \\ &= \mathbf{E}[I_{A\cap\{\tau=k-1\}}f_\tau] + \mathbf{E}[I_{A\cap\{\tau\geqslant k\}}\mathbf{E}(f_\tau|\mathscr{F}_{k-1})] \\ &= \mathbf{E}[I_{A\cap\{\tau=k-1\}}f_\tau] + \mathbf{E}[I_{A\cap\{\tau\geqslant k\}}\mathbf{E}(\mathbf{E}(f_{\overline{\tau}}|\mathscr{F}_k)|\mathscr{F}_{k-1})] \\ &\leqslant \mathbf{E}[I_{A\cap\{\tau=k-1\}}f_{k-1}] + \mathbf{E}[I_{A\cap\{\tau\geqslant k\}}\mathbf{E}(v_k|\mathscr{F}_{k-1})] \leqslant \mathbf{E}[I_A v_{k-1}].\end{aligned} \tag{15}$$

由于集合 A 的 $\mathscr{F}_{k-1}-$ 可测性, 由此可见, 对于任意设 $\tau\in\mathfrak{M}_{k-1}(\mathbf{P}-\text{a.c.})$,

$$\mathbf{E}(f_\tau|\mathscr{F}_{k-1}) \leqslant v_{k-1}. \tag{16}$$

现在证明, 对于时间 τ_{k-1}, 依 $\mathbf{P}-$ 概率 1, 有

$$\mathbf{E}(f_{\tau_{k-1}}|\mathscr{F}_{k-1}) = v_{k-1}. \tag{17}$$

(假如能证明该式, 则由 (16) 式可见, 对于 $n = k-1$, 关系式 (10) 和 (11) 也成立).

为此, 只需证明, 对于 $\tau = \tau_{k-1}$, 实际上 (15) 式到处为等式.

像 (15) 式那样开始, 并注意到根据 (5) 式的定义在集合 $\{\tau_{k-1} \geqslant k\}$ 上 $\tau = \tau_k$, 则 (根据归纳法的假设) 由 $\mathbf{E}(f_{\tau_k}|\mathscr{F}_k) = v_k(\mathbf{P}-\text{a.c.})$, 可得

$$\begin{aligned}\mathbf{E}[I_A f_{\tau_{k-1}}] &= \mathbf{E}[I_{A\cap\{\tau_{k-1}=k-1\}}f_{k-1}] + \mathbf{E}[I_{A\cap\{\tau_{k-1}\geqslant k\}}\mathbf{E}(f_{\tau_{k-1}}|\mathscr{F}_{k-1})] \\ &= \mathbf{E}[I_{A\cap\{\tau_{k-1}=k-1\}}f_{k-1}] + \mathbf{E}[I_{A\cap\{\tau_{k-1}\geqslant k\}}\mathbf{E}(f_{\tau_k}|\mathscr{F}_{k-1})] \\ &= \mathbf{E}[I_{A\cap\{\tau_{k-1}=k-1\}}f_{k-1}] + \mathbf{E}[I_{A\cap\{\tau_{k-1}\geqslant k\}}\mathbf{E}(v_k|\mathscr{F}_{k-1})] = \mathbf{E}[I_A v_{k-1}],\end{aligned}$$

其中在证明最后一个等式时, 根据 (6) 式, 有: $v_{k-1} = \max(f_{k-1}, \mathbf{E}(v_k|\mathscr{F}_{k-1}))$; 由此可见, 在集合 $\{\tau_{k-1} = k-1\}$ 上 $v_{k-1} = f_{k-1}$, 而在集合 $\{\tau_{k-1} > k-1\} = \{\tau_{k-1} \geqslant k\}$ 上 $v_{k-1} > f_{k-1}$ (因此, 在这一集合上 $v_{k-1} = \mathbf{E}(v_k|\mathscr{F}_{k-1})$).

这样, 性质 (17) 得证. 像前面已经指出的那样, 该性质连同 (16) 式就证明了欲证的关系式 (10) 和 (11).

由这些关系式可见, 对于 $\tau \in \mathfrak{M}_n(=\mathfrak{M}_n^N), (\mathbf{P}-\text{a.c.})$ 有

$$v_n = \mathbf{E}(f_{\tau_n}|\mathscr{F}_n) \geqslant \mathbf{E}(f_\tau|\mathscr{F}_n). \tag{18}$$

从而, 注意到 $v_n^N = v_n$, 由此可得

$$\mathbf{E}v_n^N = \mathbf{E}f_{\tau_n} \geqslant \sup_{\tau\in\mathfrak{M}_n^N} \mathbf{E}f_\tau = V_n^N, \tag{19}$$

而这就证明了欲证的关系式 (9) 和 (12).

性质 (13) 是如下两个性质的特殊情形: 1) 性质 (13) 是性质 (12) 当 $n=0$ 时的情形, 2) v_0^N 是常数 (因为由 (11) 式, 以及 $\sigma-$ 代数 $\mathscr{F}_0 = (\varnothing, \Omega)$ 是平凡的). 最后, 等式 (14) 当 $n = N$ 时是定义 (5) 的推论. □

5. 最优停时的 "鞅" 视角 为显示所研究的最优停止问题的 "鞅" 视角, 现在讨论序列 $v^N = (v_0^N, v_1^N, \cdots, v_N^N)$, 的递推关系式 (6), 其中 $v_N^N = f_N$ 是 "边界" 条件.

由 (6) 式可见, 对于每个 $n = 0, 1, \cdots, N-1, (\mathbf{P}-\text{a.c.})$ 有

$$v_n^N \geqslant f_n, \tag{20}$$

$$v_n^N \geqslant \mathbf{E}(v_{n+1}^N|\mathscr{F}_n). \tag{21}$$

第一个不等式在这里表示, 序列 v^N 控制序列 $f = (f_0, f_1, \cdots, f_N)$. 第二个不等式表示, 序列 v^N 是 "术语" 值为 $v_N^N = f_N$ 的上鞅. 这样, 可以说, 序列 $v^N = (v_0^N, v_1^N, \cdots, v_N^N)$, 其中由 (6) 式或 (11) 式决定于的量 v_n^N, 是上鞅的控制序列 $f = (f_0, f_1, \cdots, f_N)$.

换一种说法, 这表示序列 v^N 属于序列 $\gamma^N = (\gamma_0^N, \gamma_1^N, \cdots, \gamma_N^N)$ 类, 其中 $\gamma_n^N \geqslant f_N$, 且对于一切 $n = 0, 1, \cdots, N-1, (\mathbf{P}-\text{a.c.})$ 满足 "变分不等式"

$$\gamma_n^N \geqslant \max(f_n, \mathbf{E}(\gamma_{n+1}^N|\mathscr{F}_n)). \tag{22}$$

但是序列 v^N 具有如下补充性质: (22) 式不仅是“不严格”不等式“$\geqslant$”, 而且就是等式“=”(见 (6) 式). 根据这条性质, 就可以用如下方式从序列 $\gamma^N = (\gamma_0^N, \gamma_1^N, \cdots, \gamma_N^N)$ 类 (其中 $\gamma_N^N \geqslant f_N$) 中, 划分出序列 v^N.

定理 2 *序列 v^N 是最小上鞅的控制序列 $f = (f_0, f_1, \cdots, f_N)$.*

证明 事实上, 由于 $v_N^N = f_N$, 而 $\gamma_N^N \geqslant f_N$, 可见 $\gamma_N^N \geqslant v_N^N$. 由此以及 (22) 式和 (6) 式, ($\mathbf{P}$-a.c.) 有

$$\gamma_{N-1}^N \geqslant \max(f_{N-1}, \mathbf{E}(\gamma_N^N | \mathscr{F}_{N-1})) \geqslant \max(f_{N-1}, \mathbf{E}(v_N^N | \mathscr{F}_{N-1})) = v_{N-1}^N.$$

类似地, 对于其余 $n < N-1$, 可得 $\gamma_n^N \geqslant v_n^N$.

注 该定理得结果可以表述为如下形式: *递推方程组*

$$v_n^N = \max(f_n, \mathbf{E}(v_{n+1}^N | \mathscr{F}_n)), \quad n < N,$$

的解 $v^N = (v_0^N, v_1^N, \cdots, v_N^N)$, 其中 $v_N^N = f_N$, 是递推不等式组

$$\gamma_n^N \geqslant \max(f_n, \mathbf{E}(\gamma_{n+1}^N | \mathscr{F}_n)), \quad n < N, \tag{23}$$

一切可能解 $\gamma^N = (\gamma_0^N, \gamma_1^N, \cdots, \gamma_N^N)$ 中最小者, 其中 $\gamma_N^N \geqslant f_N$.

6. 停时集与继续观测集 定理 1 和定理 2, 不仅描绘寻找价格 $V_0^N = \sup \mathbf{E} f_\tau$ 的方法, 其中上确界在马尔可夫类时间 $\mathfrak{M}_0^N$ 上求, 但同时表明如何求最优时间 τ_0^N, 即使 $\mathbf{E} f_{\tau_0^N} = V_0^N$ 的时间.

根据 (8) 式

$$\tau_0^N = \min\{0 \leqslant k \leqslant N : v_k^N = f_k\}. \tag{24}$$

在求解关于最优停时问题时, 如下这一停时 τ_0^N 的等价描绘很重要. 设

$$D_n^N = \{\omega : v_n^N(\omega) = f_n(\omega)\} \tag{25}$$

和

$$C_n^N = \Omega \backslash D_n^N = \{\omega : v_n^N(\omega) = \mathbf{E}(v_{n+1}^N | \mathscr{F}_n)(\omega)\}.$$

显然, $D_N^N = \Omega, C_N^N = \varnothing$ 和

$$D_0^N \subseteq D_1^N \subseteq, \cdots, \subseteq D_N^N = \Omega,$$

$$C_0^N \supseteq C_1^N \supseteq, \cdots, \supseteq C_N^N = \varnothing.$$

由 (24) 和 (25) 式可见, 时间 τ_0^N可由下列形式定义

$$\tau_0^N = \min\{0 \leqslant k \leqslant N : \omega \in D_k^N\}. \tag{26}$$

自然, 称区域 D_k^N 为“停时集合”, 称区域 C_k^N 为“继续观测集合”. 这些术语的正确性有如下根据.

假设时间 $n=0$, 并将集合 Ω 划分为两个集合 D_0^N 和 C_0^N, 使 $D_0^N\bigcup C_0^N=\Omega, D_0^N\bigcap C_0^N=\varnothing$. 如果 $\omega\in D_0^N$, 则 $\tau_0^N(\omega)=0$. 换句话说, "停止" 发生在时刻 $n=0$. 而若 $\omega\in C_0^N$, 则说明对于这样的 ω, 时间 $\tau_0^N(\omega)\geqslant 1$. 对于所考虑的 $\omega\in D_0^N\bigcap C_0^N$ 的情形, 时间 $\tau_0^N(\omega)=1$. 以下各个阶段的情形类似. 在时刻 N 观测自然应该停止.

7. 例　现在举几个例子.

例 1　设序列 $f=(f_0,f_1,\cdots,f_N)$ 是鞅, 其中 $f_0=1$. 那么, 根据 §2 定理 1 的系 1, 对于一切马尔可夫时间 $\tau\in\mathfrak{M}_0^N, \mathbf{E}f_\tau=1$. 因此, 在所考虑的情形下

$$V_0^N=\sup_{\tau\in\mathfrak{M}_0^N}\mathbf{E}f_\tau=1.$$

对于一切 $1\leqslant n\leqslant N$, 函数 $v_n^N=f_n$ 和 $v_0^N=1$. 易见, $\tau_0^N=\min\{0\leqslant k\leqslant N: f_k=v_k^N\}=0$, 而对于任意 $1\leqslant n\leqslant N, \tau_n^N=n$.

于是, 鞅序列的最优停时问题的解法, 实质上是平凡的: 最优停时就是时间 $\tau_0^N(\omega)=0, \omega\in\Omega$ (任何其他时间, 如 $\tau_n^N(\omega)=n, \omega\in\Omega, 1\leqslant n\leqslant N$ 也一样).

例 2　设序列 $f=(f_0,f_1,\cdots,f_N)$ 是下鞅, 则对于任何 $\tau\in\mathfrak{M}_0^N$, 有 $\mathbf{E}f_\tau\leqslant\mathbf{E}f_N$ (§2 定理 1). 因此, 这里的最优停时 $\tau^*\equiv N$. 由于 $v_k^N=\mathbf{E}(f_N|\mathscr{F}_k)\geqslant f_k(\mathbf{P}-\text{a.c.})$, 故完全可能停时 $\tau_0^N(\omega)$ 对于某个 $\omega\in\Omega$, 小于 N. 然而, 在任何情形下停时 τ_0^N和停时 $\tau^*\equiv N$ 都是最优的. 虽然停时 $\tau^*\equiv N$ 有简单的构造, 但是停时 τ_0^N 具有一定的优越性: 它在全部可能的最优停时中是最小的, 即如果 $\widetilde{\tau}$ 也是类 $\mathfrak{M}_0^N$ 中的最优停时, 则 $\mathbf{P}\{\tau_0^N\leqslant\widetilde{\tau}\}=1$.

例 3　设序列 $f=(f_0,f_1,\cdots,f_N)$ 是上鞅. 那么, 对于一切 $0\leqslant n\leqslant N$, 有 $v_n^N=f_n$. 从而, 像鞅的情形一样, 最优停时为 $\tau_0^N=0$.

所引进的例子相当简单, 解所考虑的停时的最优性问题, 本质上并未用到在定理 1 和定理 2 中阐述的理论. 只用到如下熟知的结果: 在对马尔可夫时间作时间变换时 (见 §2), 鞅性, 下鞅性和上鞅性的不变性. 然而, 在一般情形下, 寻求价格 V_0^N, 以及最优停时 τ_0^N 的问题, 可能是相当困难的问题.

非常重要意义的是, 函数 f_n 具有如下形式的情形:

$$f_n(\omega)=f(X_n(\omega)),$$

其中 $X=(X_n)_{n\geqslant 0}$ 是某一马尔可夫链. 由第八章 §9 可见, 这时求解最优停时问题, 实质上归结为求解变分不等式, 求解瓦尔德—贝尔曼 (A. Wald – R. E. Bellman) 动态规划方程问题.

在第八章的 §9 中还将要给出, 非平凡例子, 在这些例子中, 将给出关于马尔可夫序列最优停时一系列问题的全解.

8. 练习题

1. 设 $\xi(\omega)=\sup\limits_{\alpha\in\mathfrak{U}_0}\xi_\alpha(\omega)$ 是, 在证明第 3 小节的引理时引进的随机变量. 证明满足本质上确界定义中的条件 a) 和 b). (提示: 在 $\alpha\notin\mathfrak{U}_0$ 的情形下考虑 $\mathbf{E}\max(\xi(\omega),\xi_\alpha(\omega)))$.

2. 证明, 1) 随机变量 $\xi(\omega)=\tan\widetilde{\xi}(\omega)$ (见第 3 小节引理的证明), 2) 随机变量 $\xi(\omega)$ 也满足条件 a) 和 b).

3. 设 $\xi_1,\xi_2,\cdots$ 是独立同分布随机变量序列, 其中 $\mathbf{E}|\xi_1(\omega)|<\infty$. 考虑最优停时问题 (在类 $\mathfrak{M}_1^\infty=\{\tau:1\leqslant\tau<\infty\}$ 中):

$$V^*=\sup_{\tau\in\mathfrak{M}_1^\infty}\mathbf{E}\left(\max_{i\leqslant\tau}\xi_i-c\tau\right).$$

设 $\tau^*=\inf\{n\geqslant 1:\xi_n\geqslant A^*\}$, 其中 A^* 是方程 $\mathbf{E}(\xi_1-A^*)=c$ 的唯一解.

证明, 如果 $\mathbf{E}\{\tau^*<\infty\}=1$, 则在 $\mathbf{E}(\max_{i\leqslant\tau}\xi_i(\omega)-c\tau)$ 存在的所有有限停时 τ 类中, 停时 τ^* 是最优的. 同样证明 $V^*=A^*$.

4. 假设在该题和下题中, 有

$$\mathfrak{M}_n^\infty=\{\tau:n\leqslant\tau<\infty\},$$
$$V_n^\infty=\sup_{\tau\in\mathfrak{M}_n^\infty}\mathbf{E}f_\tau,$$
$$v_n^\infty=\operatorname*{ess\,sup}_{\tau\in\mathfrak{M}_n^\infty}\mathbf{E}(f_\tau|\mathscr{F}_n),$$
$$\tau_n^\infty=\inf\{k\geqslant n:v_n^\infty=f_n\}.$$

假设

$$\mathbf{E}\sup f_n^-<\infty,$$

证明, 对于随机变量的极限

$$\widetilde{v}_n=\lim_{N\to\infty}v_n^N,$$

下列命题成立:

(a) 如果 $\tau\in\mathfrak{M}_n^\infty$ 停时, 则

$$\widetilde{v}_n\geqslant\mathbf{E}(f_\tau|\mathscr{F}_n);$$

(b) 如果 $\tau_n^\infty\in\mathfrak{M}_n^\infty$ 停时, 则

$$\widetilde{v}_n=\mathbf{E}(f_{\tau_n^\infty}|\mathscr{F}_n),$$
$$\widetilde{v}_n=v_0^\infty\left(=\operatorname*{ess\,sup}_{\tau\in\mathfrak{M}_n^\infty}\mathbf{E}(f_\tau|\mathscr{F}_n)\right).$$

5. 设 $\tau_n^\infty \in \mathfrak{M}_n^\infty$. 由上题的论断 (a) 和 (b) 证明, 停时 τ_n^∞ 按如下意义是最优的:

$$\operatorname*{ess\,sup}_{\tau \in \mathfrak{M}_n^\infty} \mathbf{E}(f_\tau|\mathscr{F}_n) = \mathbf{E}(f_{\tau_n^\infty}|\mathscr{F}_n) \quad (\mathbf{P}-\text{a.c.}),$$

而

$$\sup_{\tau \in \mathfrak{M}_n^\infty} \mathbf{E} f_\tau = \mathbf{E} f_{\tau_n^\infty},$$

即 $V_n^\infty = \mathbf{E} f_{\tau_n^\infty}$.



第八章　形成马尔可夫链的随机变量序列





现代马尔可夫过程论的源泉, 一方面是, 马尔可夫 (A. A. Марков) 有关 "联系为链" 的试验序列 (1906 年 — 1917 年) 的工作; 另一方面是, 著名布朗运动的物理现象之数学描述的尝试 (巴切利耶 [L. Bachelier], 1900 年; 爱因斯坦 [A. Einstein], 1905 年).

邓肯 (Е. Б. Дынкин)《马尔可夫过程》[21]

§1. 定义和基本性质

1. 引言 在第一章 §12 中, 对于有限概率空间, 阐述了基于对随机变量马尔可夫相依性的理解和原理 (第一章 §12 中性质 (7)), 马尔可夫相依性用来描绘, 具有无后效性系统的演变. 在这一节, 对更加一般的概率空间进行相应的研究.

马尔可夫理论的基本问题之一, 就是研究无后效系统 (随时间的发展) 的渐近性质. 非常值得注意的是, 在相当广泛的条件下, 这样系统的演变似乎 "忘掉了" 它初始的状态, 其状态 "趋于稳定", 系统进入 "平稳状态". 接下来对用 "具有可数种状态的马尔可夫链", 详细地描绘这样系统的渐近状态问题. 为此, 我们需要对 "马尔可夫链" 的状态, 按转移概率的代数性质和渐近性质进行分类.

2. 广义马尔可夫链 假设 $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 0}, \mathbf{P})$ 是固定的概率空间, 即具有另外选定附加构造的概率空间 $(\Omega, \mathscr{F}, \mathbf{P})$, 而所谓附加构造就是 $\sigma-$ 代数 $\mathscr{F}_n (n \geqslant 0)$ 的、过滤结构 (流) $(\mathscr{F}_n)_{n\geqslant 0}$, 其中 $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \cdots \subseteq \mathscr{F}$. 直观上, $\mathscr{F}_n$ 是在时间 n 之前 (包括 n) 获得的 "信息".

设 $(E, \mathscr{E})$ 是某一可测空间, 是所研究的系统取值的状态空间. 由于 "技术" 的原因 (例如, 为了使随机元素 $X_0(\omega)$ 和 $x \in E$ 的集合 $\{\omega : X_0(\omega) = x\}$ 属于 $\mathscr{F}$), 我们假设 $\sigma-$ 代数 $\mathscr{E}$ 包含 E 中的全部单点集.(关于这一假设, 亦见下面的第 6 小节).

在这样的假设下, 可测空间 $(E, \mathscr{E})$ 通常称为 (所研究系统的) 相空间或状态空间.

定义 1 (广义马尔可夫链) 设 $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 0}, \mathbf{P})$ 是固定的概率空间, 而 $(E, \mathscr{E})$ 是相空间.

定义在空间 $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 0}, \mathbf{P})$ 上、取值于 E 的 $\mathscr{F}/\mathscr{E}-$ 可测随机元 $X_n = X_n(\omega), n \geqslant 0$, 的序列 $X = (X_n)_{n\geqslant 0}$, 称为由马尔可夫相依性联系的广义随机变量序列 (马尔可夫链), 如果对于任意 $n \geqslant 0$ 和 $B \in \mathscr{E}$ 满足如下广义马尔可夫性[①]:

$$\mathbf{P}(X_{n+1} \in B | \mathscr{F}_n)(\omega) = \mathbf{P}(X_{n+1} \in B | X_n(\omega)) \quad (\mathbf{P} - \text{a.c.}). \tag{1}$$

①广义马尔可夫性(wide - sense Markov proberty) 亦称弱马尔可夫性(weak Markov proberty). —— 译者

如果 $\mathscr{F}_n^X = \sigma(X_0, X_1, \cdots, X_n)$ 是由随机变量 $X_0, X_1, \cdots, X_n$ 生成的 $\sigma-$ 代数, 则由于 $\mathscr{F}_n^X \subseteq \mathscr{F}_n$, 且 X_n 为 $\mathscr{F}_n^X-$ 可测, 故由 (1) 式得**强马尔可夫性** (或简称**马尔可夫性**):

$$\mathbf{P}(X_{n+1} \in B|\mathscr{F}_n^X)(\omega) = \mathbf{P}(X_{n+1} \in B|X_n(\omega)) \quad (\mathbf{P}-\text{a.c.}). \tag{2}$$

为直观计 (对照第一章 §12), 强马尔可夫性常表示为:

$$\mathbf{P}(X_{n+1} \in B|X_0(\omega), \cdots, X_n(\omega)) = \mathbf{P}(X_{n+1} \in B|X_n(\omega)) \ (\mathbf{P}-\text{a.c.}). \tag{3}$$

受由 (1) 式引出的强马尔可夫性 (2) 的启发, 可见在事先不考虑 $\sigma-$ 代数流 $(\mathscr{F}_n)_{n\geqslant 0}$ 的情况下, 引进马尔可夫相依性的概念较为适宜.

定义 2 (马尔可夫链) 设 $(\Omega, \mathscr{F}, \mathbf{P})$ 是概率空间, 而 $(E, \mathscr{E})$ 是相空间. 取值于 E 且 $\mathscr{F}/\mathscr{E}-$ 可测随机元 $X_n = X_n(\omega)$ 序列, $X = (X_n)_{n\geqslant 0}$, 称为**由马尔可夫相依性联系的广义随机变量序列 (马尔可夫链)**, 如果对于任意 $n \geqslant 0$ 和 $B \in \mathscr{E}$, **强马尔可夫性** (2) 成立.

注 在一开始引进的过滤概率空间上, 在此空间上根据各种不同的 "信息流" $(\mathscr{F}_n)_{n\geqslant 0}$ 研究系统的性质. 在此空间上定义的广义马尔可夫链, 在许多问题中是有效的, 例如, 可能出现这样的情况: 对于 "二维" 过程 $(X, Y) = (X_n, Y_n)_{n\geqslant 0}$, 其第一个分量 $X = (X_n)_{n\geqslant 0}$ 按定义 (2) 不是马尔可夫过程, 然而按定义 (1) 是马尔可夫过程, 其中 $\mathscr{F}_n = \mathscr{F}_n^{X,Y}, n \geqslant 0$.

在介绍**马尔可夫链**的初等理论时, 通常并不引进 $\sigma-$ 代数流 $(\mathscr{F}_n)_{n\geqslant 0}$, 而且以定义 2 作为基础. 这一章的内容就是讲马尔可夫链的初等理论.

3. 马尔可夫性 在由序列 $X = (X_n)_{n\geqslant 0}$ 描绘的系统状态的演变中, 马尔可夫性的特征是 "无后效性". 在有限空间 Ω 的情形下, 在第一章 §12 中, 表示为性质

$$\mathbf{P}(\mathrm{J}|\mathrm{GX}) = \mathbf{P}(\mathrm{J}|\mathrm{X}), \tag{4}$$

其中 J —— 将来, G —— 过去, X —— 现在. 这在第一章 §12 中曾经指出, 马尔可夫系统还具有如下性质:

$$\mathbf{P}(\mathrm{GJ}|\mathrm{X}) = \mathbf{P}(\mathrm{G}|\mathrm{X})\mathbf{P}(\mathrm{J}|\mathrm{X}), \tag{5}$$

表示在 "现在" 固定的情形下, "过去" 和 "将来" 独立.

在一般情形下, 下面的定理, (在定义 2 意义上) 给出了马尔可夫性的各种不同表述, 并指出性质 (6) 和 (7) 是性质 (4) 和 (5) 的类似, 使用了如下记号:

$$\mathscr{F}_{[0,n]}^X = \sigma(X_0, X_1, \cdots, X_n),$$
$$\mathscr{F}_{[n,\infty)}^X = \sigma(X_n, X_{n+1}, \cdots),$$
$$\mathscr{F}_{(n,\infty)}^X = \sigma(X_{n+1}, X_{n+2}, \cdots).$$

定理 1 马尔可夫性 (2), 等价于如下两条性质中的任意一条:

1) 对于 $n \geqslant 0$ 和任意 "将来的" 事件 $\mathrm{J} \in \mathscr{F}^X_{[0,\infty]}$, 有

$$\mathbf{P}(\mathrm{J}|\mathscr{F}^X_{[0,n]})(\omega) = \mathbf{P}(\mathrm{J}|X_n(\omega)) \quad (\mathbf{P}-\text{a.c.}), \tag{6}$$

2) 对于 $n \geqslant 1$ 以及任意 "将来的" 事件 $\mathrm{J} \in \mathscr{F}^X_{(n,\infty)}$ 和 "过去的" 事件 $\mathrm{G} \in \mathscr{F}^X_{[0,n-1]}$, 有

$$\mathbf{P}(\mathrm{GJ}|X_n(\omega)) = \mathbf{P}(\mathrm{G}|X_n(\omega))\mathbf{P}(\mathrm{J}|X_n(\omega)) \quad (\mathbf{P}-\text{a.c.}). \tag{7}$$

证明 (a) 我们首先证明性质 (6) 和 (7) 的等价性.

(6)⇒(7). $(\mathbf{P}-\text{a.c.})$ 有

$$\begin{aligned}
&\mathbf{P}(\mathrm{G}|X_n(\omega))\mathbf{P}(\mathrm{J}|X_n(\omega)) = \mathbf{E}(I_{\mathrm{G}}|X_n(\omega))\mathbf{E}(I_{\mathrm{J}}|X_n(\omega))\\
&= \mathbf{E}\{I_{\mathrm{G}}\mathbf{E}(I_{\mathrm{J}}|X_n(\omega)|X_n(\omega))\} = \mathbf{E}\{I_{\mathrm{G}}\mathbf{E}(I_{\mathrm{J}}|\mathscr{F}^X_{[0,n]})(\omega)|X_n(\omega)\}\\
&= \mathbf{E}\{\mathbf{E}(I_{\mathrm{G}}I_{\mathrm{J}}|\mathscr{F}^X_{[0,n]})(\omega)|X_n(\omega)\} = \mathbf{E}\{I_{\mathrm{G}}I_{\mathrm{J}}|X_n(\omega)\} = \mathbf{P}(\mathrm{GJ}|X_n(\omega)).
\end{aligned}$$

(7)⇒(6). 需要证明, 对于任意集合 $C \in \mathscr{F}^X_{[0,n]}$, 有

$$\mathbf{E}[I_C\mathbf{P}(\mathrm{J}|X_n)] = \mathbf{E}[I_C\mathbf{P}(\mathrm{J}|\mathscr{F}^X_{[0,n]})]. \tag{6'}$$

为此, 首先考虑这样集合的特殊情形, 即集合 GX, 其中 $\mathrm{G} \in \mathscr{F}^X_{[0,n-1]}$ 和 $X \in \sigma(X_n)$, 并且证明这时由 (7) 式可得 (6′) 式.

事实上,

$$\begin{aligned}
&\mathbf{E}[I_{\mathrm{GX}}\mathbf{P}(\mathrm{J}|X_n)] = \mathbf{E}[I_{\mathrm{G}}I_{\mathrm{X}}\mathbf{E}(\mathrm{J}|X_n)] = \mathbf{E}\{I_{\mathrm{X}}\mathbf{E}[I_{\mathrm{G}}\mathbf{E}(I_{\mathrm{J}}|X_n)|X_n]\}\\
&= \mathbf{E}\{I_{\mathrm{X}}\mathbf{E}(I_{\mathrm{G}}|X_n)\mathbf{E}(I_{\mathrm{J}}|X_n)\} = \mathbf{E}\{I_{\mathrm{X}}\mathbf{P}(\mathrm{G}|X_n)\mathbf{P}(\mathrm{J}|X_n)\}\\
&\overset{(7)}{=} \mathbf{E}\{I_{\mathrm{X}}\mathbf{P}(\mathrm{GJ}|X_n)\} = \mathbf{P}(\mathrm{GXJ}) = \mathbf{E}[I_{\mathrm{GX}}\mathbf{P}(\mathrm{J}|\mathscr{F}^X_{[0,n]})],
\end{aligned} \tag{8}$$

即对于形如 GX 的集合 C, (6′) 式成立, 其中 $\mathrm{G} \in \mathscr{F}^X_{[0,n-1]}$ 和 $\mathrm{X} \in \sigma(X_n)$. 根据 "单调类" 的性质 (见第二章 §2), 由此可见, 对于任意集合 $C \in \mathscr{F}^X_{[0,n]}$, 性质 (6′) 成立. 因为函数 $\mathbf{P}(\mathrm{J}|X_n)$ 为 $\mathscr{F}^X_{[0,n]}-$ 可测, 所以由 (6′) 式可见. $\mathbf{P}(\mathrm{J}|X_n)$ 是条件概率 $\mathbf{P}(\mathrm{J}|\mathscr{F}^X_{[0,n]})$ 的变式, 于是, 性质 (6) 式成立.

(b) 现在证明性质 (2) 和 (6) 等价, 因此由已证明的, 可见性质 (2) 和 (7) 等价. (6)⇒(2) 显然. 现在证明 (2)⇒(6). 仍然利用 "单调类" 的性质证明之.

式 (6) 中集合 J 是 $\sigma-$ 代数 $\mathscr{F}^X_{(n,\infty)} = \mathscr{F}^X_{[n+1,\infty)}$ 中的子集, 而 $\mathscr{F}^X_{[n+1,\infty)}$ 是代数 $\bigcup\limits_{n=1}^{\infty}\mathscr{F}^X_{[n+1,n+k]}$ 诱导的 $\sigma-$ 代数, 其中 $\mathscr{F}^X_{[n+1,n+k]} = \sigma(X_{n+1},\cdots,X_{n+k})$. 因此, 自然首先对于 $\sigma-$ 代数 $\mathscr{F}^X_{[n+1,n+k]}$ 中的集合 J 证明性质 (6).

我们用归纳法来证明. 假如 $k=1$, 则 $\mathscr{F}^X_{[n+1,n+1]} = \sigma(X_{n+1})$, 且 (6) 式恰好是 (2) 式, 故命题成立.

现在假设 (6) 式对于某个 $k\geqslant 1$ 成立, 我们证明 (6) 式对于 $k+1$ 成立.

为此取形如 $\mathrm{J}=\mathrm{J}^1+\mathrm{J}^2$ 的集合 $\mathrm{J}\in\mathscr{F}^X_{[n+1,n+k+1]}$, 其中 $\mathrm{J}^1\in\mathscr{F}^X_{[n+1,n+k]}$ 和 $\mathrm{J}^2\in\sigma(X_{n+k+1})$. 那么, 根据归纳法的假设, $(\mathbf{P}-\text{a.c.})$ 有

$$\begin{aligned}\mathbf{P}(\mathrm{J}|\mathscr{F}^X_{[0,n]})&=\mathbf{E}(I_{\mathrm{J}}|\mathscr{F}^X_{[0,n]})=\mathbf{E}(I_{\mathrm{J}^1\cap\mathrm{J}^2}|\mathscr{F}^X_{[0,n]})=\mathbf{E}[I_{\mathrm{J}^1}\mathbf{E}(I_{\mathrm{J}^2}|\mathscr{F}^X_{[0,n+k]}|\mathscr{F}^X_{[0,n]})]\\&=\mathbf{E}[I_{\mathrm{J}^1}\mathbf{E}(I_{\mathrm{J}^2}|X_{n+k})|\mathscr{F}^X_{[0,n]}]=\mathbf{E}[I_{\mathrm{J}^1}\mathbf{E}(I_{\mathrm{J}^2}|X_{n+k})|X_n]\\&=\mathbf{E}[I_{\mathrm{J}^1}\mathbf{E}(I_{\mathrm{J}^2}|\mathscr{F}_{[n,n+k]})|X_n]=\mathbf{E}[\mathbf{E}(I_{\mathrm{J}^1}I_{\mathrm{J}^2}|\mathscr{F}_{[n,n+k]})|X_n]\\&=\mathbf{E}[I_{\mathrm{J}^1}I_{\mathrm{J}^2}|X_n]=\mathbf{P}(\mathrm{J}^1\bigcap\mathrm{J}^2|X_n)=\mathbf{P}(\mathrm{J}|X_n).\end{aligned}\tag{9}$$

这样, 对形如 $\mathrm{J}=\mathrm{J}^1+\mathrm{J}^2$ 的集合 $\mathrm{J}\in\mathscr{F}^X_{[n+1,n+k+1]}$, 其中 $\mathrm{J}^1\in\mathscr{F}^X_{[n+1,n+k]}$ 和 $\mathrm{J}^2\in\sigma(X_{n+k+1})$, 已经证明了性质 (9). 由此可见 (练习题 1a), 性质 (9) 对于任意集合 $\mathrm{J}\in\mathscr{F}^X_{[n+1,n+k+1]}$ 成立. 因而 (练习题 1b), 性质 (9) 对于任意集合 $\mathrm{J}\in\bigcup\limits_{k=1}^{\infty}\mathscr{F}^X_{[n+1,n+k]}$ 代数也成立. 于是, 由此同样地可见 (练习题 1c), 性质 (9) 对于 $\sigma-$ 代数

$$\sigma\left(\bigcup_{k=1}^{\infty}\mathscr{F}^X_{[n+1,n+k]}\right)=\mathscr{F}^X_{(n,\infty)},$$

也成立. □

注 这一证明的论据是 *适当集合原理* 的应用 (首先对于 “简单” 构造的集合进行证明), 然后利用单调类的结果 (第二章 §2). 以后, 这一证明方法还将不止一次地使用 (例如, 见定理 2 和定理 3 的证明, 特别由这些定理的证明, 可以得到以上定理 1 的证明中, 引用练习题 1a, 1b, 1c 的地方).

4. 随机游动 马尔可夫链的经典例子是随机游动 $X=(X_n)_{n\geqslant 0}$, 其中

$$X_n=X_0+S_n,\quad n\geqslant 1,\tag{10}$$

而 $S_n=\xi_1+\cdots+\xi_n$, 且定义在概率空间 $(\Omega,\mathscr{F},\mathbf{P})$ 上的随机变量 $X_0,\xi_1,\xi_2,\cdots$ 相互独立.

定理 2 *假设 $\mathscr{F}_0=\sigma(X_0),\mathscr{F}_n=\sigma(X_0,\xi_1,\cdots,\xi_n),n\geqslant 1$; 而 $(\Omega,\mathscr{F},(\mathscr{F}_n)_{n\geqslant 0},\mathbf{P})$ 是过滤概率空间, 且所研究的序列 $X=(X_n)_{n\geqslant 0}$ 是该空间上的 (广义及狭义) 马尔可夫链: 对于任意 $n\geqslant 0$ 和 $B\in\mathscr{B}(\mathbb{R})$,*

$$\mathbf{P}(X_{n+1}\in B|\mathscr{F}_n)(\omega)=\mathbf{P}(X_{n+1}\in B|X_n(\omega))\ \ (\mathbf{P}-\text{a.c.}),\tag{11}$$

并且

$$\mathbf{P}(X_{n+1}\in B|X_n(\omega))=P_{n+1}(B-X_n(\omega))\ \ (\mathbf{P}-\text{a.c.}),\tag{12}$$

其中

$$P_{n+1}(A)=\mathbf{P}\{\xi_{n+1}\in A\},\tag{13}$$

而

$$B-X_n(\omega)=\{y:y+X_n(\omega)\in B\},\quad B\in\mathscr{B}(\mathbb{R}).$$

证明　我们同时证明 (11) 式和 (12) 式. 对于离散型概率空间的情形, 在第一章 §12 中曾经作过类似的证明. 初看来, 似乎这里的证明也同样简单. 然而实际上, 我们通过所作证明仍然可以体会到 "什么是证明".

考虑集合 $A \in \{X_0 \in B_0, \xi_1 \in B_1, \cdots, \xi_n \in B_n\}$, 其中 $B_i \in \mathscr{B}(\mathbb{R}), i = 0, 1, \cdots, n$. 根据条件概率 $\mathbf{P}(X_{n+1} \in B|\mathscr{F}_n)(\omega)$ 的定义 (见第二章 §7)

$$
\begin{aligned}
&\int_A \mathbf{P}(X_{n+1} \in B|\mathscr{F}_n)(\omega)\mathbf{P}(\mathrm{d}\omega) = \int_A I_{\{X_{n+1}\in B\}}(\omega)\mathbf{P}(\mathrm{d}\omega) \\
&= \mathbf{P}\{X_0 \in B_0, \xi_1 \in B_1, \cdots, \xi_n \in B_n, X_{n+1} \in B\} \\
&= \int_{B_0 \times B_1 \times \cdots \times B_n} P_{n+1}(B - [x_0 + x_1 + \cdots + x_n])P_0(\mathrm{d}x_0)\cdots P_n(\mathrm{d}x_n) \\
&= \int_A P_{n+1}(B - X_n(\omega))\mathbf{P}(\mathrm{d}\omega).
\end{aligned}
\tag{14}
$$

这样, 对于 $\mathscr{F}_n$ 中形如 $A = \{X_0 \in B_0, \xi_1 \in B_1, \cdots, \xi_n \in B_n\}$ 的集合, 有

$$
\int_A \mathbf{P}(X_{n+1} \in B|\mathscr{F}_n)(\omega)\mathbf{P}(\mathrm{d}\omega) = \int_A P_{n+1}(B - X_n(\omega))\mathbf{P}(\mathrm{d}\omega). \tag{15}
$$

显然, 上面的集合A 的集系 $\mathscr{A}_n$ 是 $\pi-$ 系 ($\Omega \in \mathscr{A}_n$ 且若 $A_1 \in \mathscr{A}_n$ 和 $A_2 \in \mathscr{A}_n$, 则 $A_1 \bigcap A_2 \in \mathscr{A}_n$; 见第二章 §2 定义 2). 其次, 以 $\mathscr{L}$ 表示一切使 (15) 式成立的集合 $A \in \mathscr{F}_n$ 的全体.

现在证明 $\mathscr{L}$ 是 $\lambda-$ 系 (见第二章 §2 定义 2). 显然 $\Omega \in \mathscr{L}$, 即满足上述定义的性质 (λ_a). 由于勒贝格积分的可加性, 由上述定义也可得性质 (λ_b). 由于勒贝格积分的单调收敛定理 (见第二章 §6), 由 $\lambda-$ 系的定义可以得到第三条性质 (λ_c).

从而, $\mathscr{L}$ 是 $\lambda-$ 系. 运用第二章 §2 定理 2 的命题 c), 可得 $\sigma(\mathscr{A}_n) \subseteq \mathscr{L}$. 注意到 $\sigma(\mathscr{A}_n) = \mathscr{F}_n$, 从而性质 (15) 对于 $A \in \mathscr{F}_n$ 集合也成立.

这样, 注意到 $P_{n+1}(B - X_n(\omega))$ 作为 ω 的函数 $\mathscr{F}_n-$ 可测 (练习题 2), 故 (根据条件概率的定义) 由 (15), 可见 $P_{n+1}(B - X_n(\omega))$ 是条件概率 $\mathbf{P}(X_{n+1} \in B|\mathscr{F}_n)(\omega)$ 的变式. 最后, 由条件数学期望的 "望远性" 性质 (见第二章 §7 性质 $\mathbf{H}^*$) 可见, ($\mathbf{P}$ – a.c.) 有

$$
\begin{aligned}
\mathbf{P}(X_{n+1} \in B|X_n)(\omega) &= \mathbf{E}[I_{\{X_{n+1}\in B\}}|X_n](\omega) = \mathbf{E}\{\mathbf{E}[(I_{\{X_{n+1}\in B\}}|\mathscr{F}_n)|X_n]\}(\omega) \\
&= \mathbf{E}[P_{n+1}(B - X_n)|X_n(\omega)] = P_{n+1}(B - X_n(\omega)).
\end{aligned}
\tag{16}
$$

于是, 两条性质 (11) 和 (12) 得证. □

注　性质 (11) 和 (12) 的正确性, 或许也可以直接由第二章 §2 引理 3 推出 (练习题 3). 我们对这些 "几乎显然" 的性质作了详细的证明, 在一定意义上是为了演示类似命题的证明技术, 证明基于适当集合原理以及单调类的有关结果.

5. 广义马尔可夫性 考虑马尔可夫性 (1). 如果 $(E,\mathscr{E})$ 是博雷尔空间, 则由第二章 §7 定理 5 可见, 对于每一个 $n \geqslant 0$, 存在正则条件分布 $P_{n+1}(x;B)$, 使 ($\mathbf{P}$ – a.c.) 有

$$\mathbf{P}(X_{n+1} \in B|X_n(\omega)) = P_{n+1}(X_n(\omega);B), \tag{17}$$

其中 $P_{n+1}(x;B), B \in \mathscr{E}, x \in E$, 具有如下性质 (见第二章 §7 定义 7):

(a) 对于每一个 x , 集函数 $P_{n+1}(X_n;\cdot)$ 是空间 $(E,\mathscr{E})$ 上的测度;

(b) 对于每一个 $B \in \mathscr{E}$, 函数 $P_{n+1}(\cdot;B)$ 为 $\mathscr{E}-$ 可测.

函数 $P_n = P_n(x;B), n \geqslant 1$, 称做转移函数 (亦称马尔可夫核).

特别重要的是下面的情形: 这些转移函数全相等 $P_1 = P_2 = \cdots$, 确切地说, 对这些条件概率 $\mathbf{P}(X_{n+1} \in B|X_n(\omega))$, $n \geqslant 0$, 存在同一个正则条件分布的变式 $P(x;B)$, 使 ($\mathbf{P}$ – a.c.) 对于一切 $n \geqslant 0, B \in \mathscr{E}$, 有

$$\mathbf{P}(X_{n+1} \in B|X_n(\omega)) = P(X_n(\omega);B), \tag{18}$$

如果变式 $P = P(x;B)$ 存在 (那么可以认为对于一切 $n \geqslant 0, P_n = P$), 则马尔可夫链称做 (对时间) 齐次的, 其转移函数为 $P(x;B), x \in E, B \in \mathscr{E}$.

马尔可夫链的齐次性直观意义很明显: 相应系统的运动在如下意义上均匀地进行, 控制系统变化的概率机制在所有时间 $n \geqslant 0$ 都保持不变(在动态系统理论中, 这种性质等同于保守性).

除转移概率 $P_1, P_2, \cdots$ 之外, 而对于齐次马尔可夫链, 除转移概率 P 之外, 马尔可夫链的重要特征是初始分布 $\pi = \pi(B), B \in \mathscr{E}$, 即由等式 $\pi(B) = \mathbf{P}\{X_0 \in B\}, B \in \mathscr{E}$, 所决定的概率分布.

数组 $(\pi, P_1, P_2, \cdots)$ 完全决定序列 $X = (X_n)_{n\geqslant 0}$ 的概率性质, 因为该序列的一切有限维分布决定于如下公式:

$$\mathbf{P}\{X_0 \in B\} = \pi(B), \quad B \in \mathscr{E},$$

而对于任意 $n \geqslant 1$ 和 $B \in \mathscr{B}(E^{n+1})(= \mathscr{E}^{n+1} = \mathscr{E} \otimes \cdots \otimes \mathscr{E}(n+1)$ 次), 有

$$\begin{aligned}&\mathbf{P}\{(X_0, X_1, \cdots, X_n) \in B\}\\ &= \int_{E\times\cdots\times E} I_B(x_0, x_1, \cdots, x_n)\pi(\mathrm{d}x_0)P_1(x_0, \mathrm{d}x_1)\cdots P_n(x_{n-1}, \mathrm{d}x_n).\end{aligned} \tag{19}$$

事实上, 首先考虑形如 $B = B_0 \times \cdots \times B_n$ 的集合 B. 那么, 当 $n = 1$ 时, 由全概

率公式 (见第二章 §7(5) 式)

$$
\begin{aligned}
&\mathbf{P}\{X_0 \in B_0, X_1 \in B_1\}\\
&= \int_\Omega I_{\{X_0\in B_0\}(\omega)}(\omega)\mathbf{P}(X_1 \in B_1|X_0(\omega))\mathbf{P}(\mathrm{d}\omega)\\
&= \int_\Omega I_{\{X_0\in B_0\}(\omega)}(\omega)P_1(B_1; X_0(\omega))\mathbf{P}(\mathrm{d}\omega)\\
&= \int_E I_{B_0}(x_0)P_1(B_1; x_0)\pi(\mathrm{d}x_0)\\
&= \int_{E\times E} I_{B_0\times B_1}(x_0, x_1)P_1(\mathrm{d}x_1; x_0)\pi(\mathrm{d}x_0).
\end{aligned}
$$

下面用归纳法证明:

$$
\begin{aligned}
&\mathbf{P}\{X_0 \in B_0, X_1 \in B_1, \cdots, X_n \in B_n\}\\
&= \int_\Omega I_{\{X_0\in B_0,\cdots,X_{n-1}\in B_{n-1}\}}(\omega)\mathbf{P}(X_n \in B_n|X_0(\omega), \cdots, X_{n-1}(\omega))\mathbf{P}(\mathrm{d}\omega)\\
&= \int_\Omega I_{\{X_0\in B_0,\cdots,X_{n-1}\in B_{n-1}\}}(\omega)\mathbf{P}(X_n \in B_n|X_{n-1}(\omega))\mathbf{P}(\mathrm{d}\omega)\\
&= \int_\Omega I_{\{X_0\in B_0,\cdots,X_{n-1}\in B_{n-1}\}}(\omega)P_n(B_n; X_{n-1}(\omega))\mathbf{P}(\mathrm{d}\omega)\\
&= \int_{E\times\cdots\times E} I_{B_0\times B_1\times\cdots\times B_{n-1}}(x_0, x_1, \cdots, x_{n-1})\\
&\quad\times P_n(B_n; x_{n-1})\mathbf{P}\{X_0 \in \mathrm{d}x_1, \cdots, X_{n-1} \in \mathrm{d}x_n\}\\
&= \int_{E\times\cdots\times E} I_{B_0\times B_1\times\cdots\times B_{n-1}\times B_n}(x_0, x_1, \cdots, x_{n-1}, x_n)\\
&\quad\times P_n(\mathrm{d}x_n; x_{n-1})P_{n-1}(\mathrm{d}x_{n-1}; x_{n-2})\cdots P_1(\mathrm{d}x_1; x_0)\pi(\mathrm{d}x_0).
\end{aligned}
$$

这样, 在集合 B 为 $B = B_0 \times \cdots \times B_n$ 情形下, 所得结果恰好为 (19) 式. 对于集合 $B \in \mathscr{B}(E^{n+1})$ 的一般情形, 可以用与定理 2 证明中类似地方的方法实现.

基于单调类的结果 (见第二章 §2), 由性质 (19) 可以导出 (练习题 4), 对于任意有界 $\mathscr{B}(E^{n+1})-$ 可测函数 $h = h(x_0, x_1, \cdots, x_n)$, 有

$$
\begin{aligned}
&\mathbf{E}h(X_0, X_1, \cdots, X_n)\\
&= \int_{E^{n+1}} h(x_0, x_1, \cdots, x_n)\pi(\mathrm{d}x_0)P_1(\mathrm{d}x_1; x_0)\cdots P_n(\mathrm{d}x_n; x_{n-1}).
\end{aligned}
\tag{20}
$$

6. 数组 $(\pi, P_1, P_2, \cdots)$ **决定的马尔可夫链** 对于 (广义或狭义) 马尔可夫链, 利用 (19) 式, 根据其初始分布 $\pi = \pi(B)$, 其中 $\pi(B) = \mathbf{P}\{X_0 \in B\}, B \in \mathscr{E}$, 以及转移概率 $P_n(x, B), n \geqslant 1, x \in E, B \in \mathscr{E}$, 就可以完整地建立任意随机变量组 $X_1, X_2, \cdots, X_n (n \geqslant 1)$ 的分布律 $\mathrm{Law}(X_0, X_1, \cdots, X_n)$.

我们现在以完全给定的数组 $(\pi, P_1, P_2, \cdots)$ 为基础, 调换关于马尔可夫链的定义全部观点. 从数组 $(\pi, P_1, P_2, \cdots)$ 出发, 把其中 π 的概率意义理解为系统初始状态的

概率分布, 而把满足第 5 小节中定义的性质 (a) 和 (b) 的函数, $P_{n+1} = P_{n+1}(x, B)$, $n \geqslant 0$, 当作转移概率, 即系统在时间 n 处于状态 x, 在时间 $n+1$ 处于集合 $B \in \mathscr{E}$ 的概率. 自然, 假如给定的是数组 $(\pi, P_1, P_2, \cdots)$, 就产生一个问题: 它一般是否对应着某一马尔可夫链, 而且以给定的 π 为其初始分布, 以给定的函数 $P_1, P_2, \cdots$ 为其转移概率?

对这一问题的答案是 "正确". 实际上, 至少对于的 $E = \mathbb{R}^d$ 情形 (见第二章 §9 的定理 1 及其系 3), 答案包含在柯尔莫戈洛夫定理中; 对于任意可测空间 $(E, \mathscr{E})$ 的情形. 包含在 I. 图尔恰 (I. Tulcea) 定理中 (见第二章 §9 的定理 2).

按照这些定理证明, 我们首先定义可测空间 $(\Omega, \mathscr{F})$: 设 $(\Omega, \mathscr{F}) = (E^\infty, \mathscr{B}(E^\infty))$, 其中 $E^\infty = E \times E \times \cdots, \mathscr{B}(E^\infty) = \mathscr{E} \otimes \mathscr{E} \otimes \cdots$; 换句话说, 把 "点" $\omega = (x_0, x_1, \cdots)$ 视为基本事件, 其中 $x_i \in E$.

设 $\mathscr{F}_n = \sigma(x_0, x_1, \cdots, x_n)$, 定义流 $(\mathscr{F}_n)_{n \geqslant 0}$. 我们以 "典则" 方式定义 $X_n(\omega) = x_n$ 的值: 若 $\omega = (x_0, x_1, \cdots)$, 则设 $X_n(\omega) = x_n$.

I. 图尔恰定理: 对于任意可测空间 $(E, \mathscr{E})$ (其中也包括所考虑的相空间), 在 $(\Omega, \mathscr{F})$ 上存在概率测度 $\mathbf{P}_\pi$, 使

$$\mathbf{P}_\pi\{X_0 \in B\} = \pi(B), \quad B \in \mathscr{E}, \tag{21}$$

而对于一切 $n \geqslant 1$, 其有限维分布为

$$\begin{aligned} &\mathbf{P}_\pi\{(X_0, X_1, \cdots, X_n) \in B\} \\ &= \int_E \pi(\mathrm{d}x_0) \int_E P_1(x_0; \mathrm{d}x_1) \cdots \int_E I_B(x_0, \cdots, x_n) P_n(x_{n-1}; \mathrm{d}x_n). \end{aligned} \tag{22}$$

定理 3 关于 (在 I. 图尔恰定理中) 引进的测度 $\mathbf{P}_\pi$, 典则给定的随机变量序列 $X = (X_n)_{n \geqslant 0}$ (在定义 2 的意义上) 是马尔可夫序列.

证明 需要证明对于 $n \geqslant 0, B \in \mathscr{E}$, ($\mathbf{P}_\pi$ – a.c.) 有

$$\mathbf{P}_\pi(X_{n+1} \in B | \mathscr{F}_n)(\omega) = \mathbf{P}_\pi(X_{n+1} \in B | X_n(\omega)), \tag{23}$$

而且, 这时对于 $n \geqslant 0$, ($\mathbf{P}$ – a.c.) 有

$$\mathbf{P}_\pi(X_{n+1} \in B | X_n(\omega)) = P_n(X_n(\omega); B). \tag{24}$$

仍然利用适当集合原理和单调类的结果 (第二章 §2) 进行证明.

像前面的作法一样, 作为适当集合, 考虑 "简单" 构造的集合 $A \in \mathscr{F}_n$, 其中 A 具有如下形式:

$$A = \{\omega : X_0(\omega) \in B_0, \cdots, X_n(\omega) \in B_n\},$$

而 $B_i \in \mathscr{E}, i = 0, 1, \cdots, n$, 并设 $B \in \mathscr{E}$.

那么, 由于测度 $\mathbf{P}_\pi$ 的构造 (见 (22) 式), 有

$$\int_A I_{\{X_{n+1}\in B\}}(\omega)\mathbf{P}_\pi(\mathrm{d}\omega) = \mathbf{P}_\pi\{X_0\in B_0,\cdots,X_n\in B_n,X_{n+1}\in B\}$$

$$= \int_{B_0}\pi(\mathrm{d}x_0)\int_{B_1}P_1(x_0;\mathrm{d}x_1)\cdots\int_{B_n}P_n(x_{n-1};\mathrm{d}x_n)\int_B P_{n+1}(x_n;\mathrm{d}x_{n+1})$$

$$= \int_B P_{n+1}(X_n(\omega);B)\mathbf{P}_\pi(\mathrm{d}\omega). \tag{25}$$

现在通过像定理 2 的证明类似的讨论 (对于集合 $A\in\mathscr{F}_n$, 见性质 (15) 的证明), 可见这里性质 (25) 对于集合 $A\in\mathscr{F}_n$ 成立, 即对于形如 $A=\{\omega:(X_0(\omega),\cdots,X_n(\omega))\in C\}$ 的集合 $A\in\mathscr{F}_n$ 成立, 其中 $C\in\mathscr{B}(E^{n+1})$.

由于条件概率的定义 (第二章 §7), 可见

$$\int_A I_{\{X_{n+1}\in B\}}(\omega)\mathbf{P}_\pi(\mathrm{d}\omega) = \int_A \mathbf{P}_\pi(X_{n+1}\in B|\mathscr{F}_n)(\omega)\mathbf{P}_\pi(\mathrm{d}\omega), \tag{26}$$

而且函数 $P_{n+1}(x_n(\omega);B)$ 为 $\mathscr{F}_n-$ 可测, 故由 (25) 式和条件数学期望的 "望远" 性质 (见第二章 §7 第 4 小节性质 $\mathbf{H}^*$) 可见, 关系式 (23) 和 (24) 成立. □

7. 马尔可夫链 (族) 上一节证明了, 与每一个给定的数组 $(\pi,P_1,P_2,\cdots)$, 有一个马尔可夫链与之相联系 (为了直观, 记作 $X^\pi=(X_n,\mathbf{P}_\pi)_{n\geqslant 0}$), 并且以 π 为其*初始分布*, 而以 $P_1,P_2,\cdots$ 为其*转移概率* (即满足性质 (21), (23) 和 (24) 的链).

在初始时间 $n=0$, 根据分布 π 随机地 "抽签决定" 初始状态的值. 例如, 结果 X_0 的值等于 x, 则系统在下一时间按照分布 $P_1(\cdot;x)$, 系统自该状态转移到某个状态 x_1, 等等.

这样, 初始分布 π 的作用仅仅表现在时间 $n=0$, 而后系统的发展就决定于转移概率 $P_1,P_2,\cdots$ 因此, 假如对于两个初始分布 π_1 和 π_2, 相应 "抽签" 的结果为同一状态 x, 那么系统的发展 (在概率意义上) 将是一样的, 都仅决定于转移概率 $P_1,P_2,\cdots$ 亦可将这一情况表示如下.

仍以 $(E,\mathscr{E})$ 表示相空间. 假设以 $\mathbf{P}_x$ 表示对应于如下情况的分布 $\mathbf{P}_\pi$: 分布 π 集中在点 x, 而 $\pi(\mathrm{d}y)=\delta_x(\mathrm{d}y)$, 即 $\pi(\{x\})=1$, 其中 $\{x\}$ 是属于 $\sigma-$ 代数 $\mathscr{E}$ 的单点集.

那么, 由性质 (22) 可见 (练习题 4), 对于每一个 $A\in\mathscr{B}(E^\infty)$ 和 $x\in E$, (对于每一个 π) 概率 $\mathbf{P}_x(A)$ 是条件概率 $\mathbf{P}_\pi(A|X_0=x)$ 的变式, 即 $\mathbf{P}_\pi$ - a.c. 有

$$\mathbf{P}_\pi(A|X_0=x)=P_x(A). \tag{27}$$

对于每一个 $x\in E$, 概率 $\mathbf{P}_x(\cdot)$ 完全决定于转移概率 $(P_1,P_2,\cdots)$ 组.

于是, 假如基本着眼点是系统对转移概率 $(P_1,P_2,\cdots)$ 的依赖关系, 则只需利用概率 $\mathbf{P}_x(\cdot), x\in E$, 而且如果需要得到概率 $\mathbf{P}_\pi(\cdot)$, 则只需进行简单的积分运算:

$$\mathbf{P}_\pi(A)=\int_E \mathbf{P}_x(A)\pi(\mathrm{d}x),\quad A\in\mathscr{B}(E^\infty). \tag{28}$$

这些思路导致如下情形: “马尔可夫过程的一般理论” 中 (见 [21]), 对于这里研究的离散时间的情形, 认为*基本研究对象*并不是哪个具体的马尔可夫链, 而是马尔可夫链*族* $X^x = (X_n, \mathbf{P}_x)_{n\geqslant 0}, x \in E$. (不过, 通常不提 “马尔可夫链族”, 仍然简称 “马尔可夫链”, 并且通常使用记号 “ $X = (X_n, \mathscr{F}_n, \mathbf{P}_x)$”, 而不是 “ $X^x = (X_n, \mathbf{P}_x)_{n\geqslant 0}, x \in E$”.)

需要强调, 这些论述都假设, 是用 “典则” 方法建立的链: 以 $(E^\infty, \mathscr{E}^\infty)$ 做 $(\Omega, \mathscr{F})$, 其中 $\mathscr{E}^\infty = \mathscr{E} \otimes \mathscr{E} \otimes \cdots$ 而若 $\omega = (x_0, x_1, \cdots)$, 则一切随机变量 $X_n(\omega)$ 定义为 $X_n(\omega) = x_n$. 这样, 在 $X^x = (X_n, \mathbf{P}_x)$ 中只有 $\mathbf{P}_x$ 依赖于 x, 而不假设 X_n 本身的值对 x 的任何特别的依赖条件, 在这种情况下, 按测度 $\mathbf{P}_x$ 轨道 $(x_n)_{n\geqslant 0}$ 自动从点 x “开始”, 即 $\mathbf{P}_x\{X_0 = x\} = 1$.

8. 柯尔莫戈洛夫 – 查普曼方程 (任意状态空间) 对于*有限*马尔可夫链的情形 (第一章 §12), 主要注意力放在, 利用转移概率 $p_{ij}^{(n)} = \mathbf{P}(X_n = j|X_0 = i)$ 的研究, 分析这种链的行为和性质方面, 证明了 $p_{ij}^{(n)}$ 满足*柯尔莫戈洛夫 – 查普曼* (D. G. Chapman) *方程* (见第一章 §12 的 (13) 式), 并且由此同样的可以得到*柯尔莫戈洛夫前向方程和后向方程* (第一章 §12 的 (15) 和 (16) 式).

我们现在讨论, 在具有*任意*相空间 $(E, \mathscr{E})$ 的情形下, 柯尔莫戈洛夫 – 查普曼方程正确性的问题, 但是只局限于*齐次*链的情形, 即 $P_1 = P_2 = \cdots = P$ 的情形.

在这种情形下, 由 (22) 式, 有

$$\begin{aligned}&\mathbf{P}_\pi\{(X_0, X_1, \cdots, X_n) \in B\}\\&= \int_E \pi(\mathrm{d}x_0) \int_E P(x_0; \mathrm{d}x_1) \cdots \int_E I_B(x_0, x_1, \cdots, x_n) P(x_{n-1}; \mathrm{d}x_n).\end{aligned} \tag{29}$$

特别, 如果 $n = 2$, 则

$$\mathbf{P}_\pi\{X_0 \in B_0, X_2 \in B_2\} = \int_{B_0} \int_E P(x_1; B_2) P(x_0; \mathrm{d}x_1) \pi(\mathrm{d}x_0). \tag{30}$$

由此, 根据拉东 – 尼科迪姆定理 (第二章 §6), 以及条件概率的定义, (π – a.c.) 有

$$\mathbf{P}_\pi\{X_2 \in B_2|X_0 = x\} = \int_E P(x; \mathrm{d}x_1) P(x_1; B_2). \tag{31}$$

现在注意到, 由于 (27) 式, 可见

$$\mathbf{P}_\pi\{X_2 \in B_2|X_0 = x\} = \mathbf{P}_x\{X_2 \in B_2\} \ \ (\pi - \text{a.c.}),$$

其中概率 $\mathbf{P}_x\{X_2 \in B_2\}$ 有简单的含义: 系统由在时间 $n = 0$ 状态 x, 于时间 $n = 2$ 转移到状态集合 B_2 的概率, 即经*两步*转移的概率.

记 $P^{(n)}(x; B_n) = \mathbf{P}_x\{X_n \in B_n\}$ 是经 n 步转移的概率. 那么, 由于所考虑的链的齐性, 有 $P^{(1)}(x; B_1) = P(x; B_1)$, 从而由 (31) 式, 可见 ($\pi$ – a.c.) 有

$$P^{(2)}(x; B) = \int_E P^{(1)}(x; \mathrm{d}x_1) P^{(1)}(x_1; B), \tag{32}$$

其中 $B \in \mathscr{E}$.

类似地可以证明 (练习题 5), 对于任意 $n \geqslant 0, m \geqslant 0$, ($\pi-$ a.c.) 有

$$P^{(n+m)}(x; B) = \int_E P^{(n)}(x; \mathrm{d}y) P^{(m)}(y; B). \tag{33}$$

该式就是著名的

柯尔莫戈洛夫 - 查普曼方程,

其直观含义十分清楚: 为计算系统 "自点 $x \in E$ 经过 $m+n$ 步转移到集合 $B \in \mathscr{E}$ 中的概率 $P^{(n+m)}(x; B)$", 需要将 "自点 x 经过 n 步转移到点 $y \in E$ 的 '无穷小' 区域 $\mathrm{d}y$ 概率" $P^{(n)}(x; \mathrm{d}y)$, 乘以 "自点 y 经过 m 步转移到集合 B 的概率", 然后对可能的 "中间点" y 积分.

柯尔莫戈洛夫 - 查普曼方程 (33), 把不同转移次数的概率相联系. 应该注意, 该式仅仅是精确到 "$\pi-$ 几乎处处" 成立. 特别, 由此可见, (33) 式并不是对于一切 $x \in E$ 都成立的关系式. 不应当觉得这是怪异的, 因为我们在前面不止一次地处理选择条件概率各种不同变式 (异说) 的问题. 一般不能指望, 所研究的性质 "对于这些变式 (关于 x) 恒成立", 而事实上 "仅仅 $\pi-$ 几乎必然 (处处) 成立".

然而, 可以指出这样一些变式, 使柯尔莫戈洛夫 - 查普曼方程 (33) 已经对于一切 $x \in E$ 成立.

这由如下命题 (见练习题 6) 可以得到.

设 "转移概率" $P^{(n)}(x; B)$ 定义为:

$$P^{(1)}(x; B) = P(x; B),$$

而对于 $n > 1$,

$$P^{(n)}(x; B) = \int_E P(x; \mathrm{d}y) P^{(n-1)}(y; B).$$

那么,

(i) 对每一个给定的 $x, P^{(n)}(x; B), n \geqslant 1$, 是 $\mathscr{E}$ 上的正则条件概率;

(ii) $P^{(n)}(x; B)$ 等于 $\mathbf{P}_x\{X_n \in B\}$, 从而 $P^{(n)}(x; B)$ ($\pi-$ a.c.) 是条件概率 $\mathbf{P}_\pi(X_n \in B | X_0 = x)$ 的变式;

(iii) 对于这样定义的函数 $P^{(n)}(x; B), n \geqslant 1$, 柯尔莫戈洛夫 - 查普曼方程 (33) 对于一切 $x \in E$ 恒成立.

9. 练习题

1. 证明在定理 1 的证明中提到的问题 1a, 1b, 1c 的命题,
2. 证明在定理 2 中的函数 $P_{n+1}(B - X_n(\omega))$ 对 ω 是 $\mathscr{F}_n-$ 可测的.
3. 证明第二章 §2 引理 3 的命题中性质 (11) 和 (12).
4. 证明性质 (20) 和 (27).

5. 证明关系式 (33).

6. 证明第 8 小节中的命题 (i), (ii), (iii).

7. 问由马尔可夫性 (3) 是否可以导出如下性质:

$$\mathbf{P}(X_{n+1} \in B | X_0 \in B_0, X_1 \in B_1, \cdots, X_n \in B_n) = \mathbf{P}(X_{n+1} \in B | X_n \in B_n),$$

其中 $B_0, B_1, \cdots, B_n$ 和 B 是属于 $\mathscr{E}$ 的集合, 而 $\mathbf{P}\{X_0 \in B_0, X_1 \in B_1, \cdots, X_n \in B_n\} > 0$.

§2. 推广马尔可夫性和强马尔可夫性

1. 推广马尔可夫性 这一节主要研究齐次马尔可夫链族 $X^x = (X_n, P_x)_{n \geqslant 0}, x \in E$, 而链是 "典则地" 定义在坐标空间 $(\Omega, \mathscr{F}) = (E^\infty, \mathscr{E}^\infty)$ 上并且决定于转移函数 $P(x; B), x \in E, B \in \mathscr{E}$.

在 $(\Omega, \mathscr{F})$ 上定义推移算子 $\theta_n : \Omega \to \Omega$ (对照第五章 §1), 对状态 $\omega = (x_0, x_1, \cdots)$, 设

$$\theta_n(\omega) = (x_n, x_{n+1}, \cdots).$$

如果 $H = H(\omega)$ 是 $\mathscr{F}-$ 可测函数, 则以 $H \circ \theta_n$ 表示由等式

$$(H \circ \theta_n)(\omega) = H(\theta_n(\omega)) \tag{1}$$

定义的函数 $(H \circ \theta_n)(\omega)$. 从而, 如果 $\omega = (x_0, x_1, \cdots)$ 和 $H = H(x_0, x_1, \cdots)$, 则

$$(H \circ \theta_n)(x_0, x_1, \cdots) = H(x_n, x_{n+1}, \cdots).$$

下面的定理实质上是 §1 中的 (6) 应用于现在考虑的齐次马尔可夫链族命题的重新表述.

定理 1 设 $X^x = (X_n, \mathbf{P}_x)_{n \geqslant 0}, x \in E$是由转移函数 $P(x; B), x \in E, B \in \mathscr{E}$ 生成的齐次马尔可夫链族. 假设对于 $B \in \mathscr{B}(E^{n+1})$ 和 $n \geqslant 0$, 由 §1 中的公式 (22) 由 P 决定测度 P_x 的值 $\mathbf{P}_x\{(X_0, X_1, \cdots, X_n) \in B\}$, 其中 $\pi(\mathrm{d}y) = \delta_{(x)}(\mathrm{d}y)$, 而 $P_1 = P_2 = \cdots = P$.

那么, 对于任意初始分布 π, 任何 $n \geqslant 0$ 以及任意有界 (或非负) $\mathscr{F}-$ 可测函数 $H = H(\omega)$, 有如下的推广马尔可夫性[①]:

$$\mathbf{E}_\pi(H \circ \theta_n | \mathscr{F}_n^X) = \mathbf{E}_{X_n(\omega)} H \quad (\mathbf{P}_\pi - \text{a.c.}). \tag{2}$$

注 虽然所使用记号自身 "无须解释", 然而我们还是指出 $\mathbf{E}_\pi$ 是按测度

$$\mathbf{P}_\pi(\cdot) = \int_E \mathbf{P}_x(\cdot) \pi(\mathrm{d}x)$$

①这里将要介绍的广义马尔可夫性 (обобщённое марковское свойство) 与一般广义马尔可夫性有所不同, 因此我们将其译为 "推广马尔可夫性", 以示区别. —— 译者

求平均, 而 $\mathbf{E}_{X_n(\omega)}H$ 应该按如下的方式理解: 求数学期望 $\mathbf{E}_xH$ (即按测度 $\mathbf{P}_x$ 求对 H 平均), 而然后向此式 (记作 $\psi(x)$) 中 x 的位置"代入"随机变量 $X_n(\omega)$, 即 $\mathbf{E}_{X_n(\omega)}H = \psi(X_n(\omega))$. (注意, $\mathbf{E}_xH$ 是 x 的 $\mathscr{E}-$ 可测函数 (练习题 1), 故 $\mathbf{E}_{X_n(\omega)}H$ 是随机变量, 即 $\mathscr{F}/\mathscr{E}-$ 可测函数). 定理的证明仍然利用适当集合与函数原理, 然后运用单调类的结果.

为证明性质 (2), 我们需要验证对属于 $\mathscr{F}_n^X = \sigma(x_0, x_1, \cdots, x_n)$ 的任意集合 A, 有

$$\int_A (H\circ\theta_n)(\omega)\mathbf{P}_\pi(\mathrm{d}\omega) = \int_A (\mathbf{E}_{X_n(\omega)}H)\mathbf{P}_\pi(\mathrm{d}\omega), \tag{3}$$

或者对于更加紧凑的形式, 有

$$\mathbf{E}_\pi(H\circ\theta_n; A) = \mathbf{E}_\pi(\mathbf{E}_{X_n}H; A), \tag{4}$$

其中 $\mathbf{E}_\pi(\xi; A)$ 表示 $\mathbf{E}_\pi(\xi I_A)$ 适当集合与函数原理 (见第二章 §6 第 2 小节).

根据适当集合原理, 我们考虑形如 $A = \{\omega : X_0 \in B_0, \cdots, X_n \in B_n\}(B_i \in \mathscr{E}_i)$ 的"简单"构成集合 A; 考虑函数 $H = H(x_0, x_1, \cdots, x_m)(m \geqslant 0)$ (更确切地说, 设 H 是 $\mathscr{F}_n^X-$ 可测函数). 那么, 性质 (4) 有如下形式:

$$\mathbf{E}_\pi[H(X_n, X_{n+1}, \cdots, X_{n+m}); A] = \mathbf{E}_\pi[\mathbf{E}_{X_n}H(X_0, X_1, \cdots, X_m); A]. \tag{5}$$

利用 §1 中的表示 (22), 可得

$$\begin{aligned}
&\mathbf{E}_\pi[H(X_n, X_{n+1}, \cdots, X_{n+m}); A] = \mathbf{E}_\pi[I_A(X_0, \cdots, X_n)H(X_n, \cdots, X_{n+m})]\\
&= \int_{E^{n+m+1}} I_A(x_0, \cdots, x_n)H(x_n, \cdots, x_{n+m})\pi(\mathrm{d}x_0)P(x_0; \mathrm{d}x_1)\cdots P(x_{n+m-1}; \mathrm{d}x_{n+m})\\
&= \int_{E^{n+1}} I_A(x_0, \cdots, x_n)\pi(\mathrm{d}x_0)P(x_0; \mathrm{d}x_1)\cdots P(x_{n-1}; \mathrm{d}x_n)\\
&\qquad\qquad \times \left[\int_{E^m} H(x_n, \cdots, x_{n+m})P(x_n; \mathrm{d}x_{n+1})\cdots P(x_{n+m-1}; \mathrm{d}x_{n+m})\right]\\
&= \int_{E^{n+1}} I_A(x_0, \cdots, x_n)\pi(\mathrm{d}x_0)P(x_0; \mathrm{d}x_1)\cdots P(x_{n-1}; \mathrm{d}x_n)\\
&\qquad\qquad \times \left[\int_{E^m} H(x_0, \cdots, x_m)\mathbf{P}_x(\mathrm{d}x_1, \cdots, \mathrm{d}x_m)\right]\\
&= \mathbf{E}_\pi[\mathbf{E}_{X_n}H(X_0, X_1, \cdots, X_m); A],
\end{aligned}$$

其中 $\mathbf{P}_x(\mathrm{d}x_1, \cdots, \mathrm{d}x_m) = P(x_0; \mathrm{d}x_1)P(x_1; \mathrm{d}x_2)\cdots P(x_{m-1}; \mathrm{d}x_m)$.

这样, 对形如 $A = \{\omega : x_0 \in B_0, \cdots, x_n \in B_n\}$ 的集合 A, 和函数 $H = H(x_0, x_1, \cdots, x_m)$, 性质 (5) 得证. 集合 $A \in \mathscr{F}_n^X$ 的一般情形 (对于固定的 m) 的证明, 同 §1 中定理 2 的证明是一样的.

只剩下证明, 已证明的性质对于一切 $\mathscr{F}(= \mathscr{E}^\infty)-$ 可测有界函数 $H = H(x_0, x_1, \cdots)$ 也仍然成立.

为证明这一事实, 只需证明, 如果 $A \in \mathscr{F}_n^X$, 则对于函数

$$\mathbf{E}_\pi[H(X_n, X_{n+1}, \cdots); A] = \mathbf{E}_\pi[\mathbf{E}_{X_n} H(X_0, X_1, \cdots); A], \tag{6}$$

性质 (5) 成立.

为了利用适当集合与函数原理 (第二章 §2), 以 $\mathscr{H}$ 表示满足性质 (5) 的一切有界 $\mathscr{F}-$ 可测函数 $H = H(x_0, x_1, \cdots)$ 的全体.

设 J 是形如 $I_m = \{\omega : x_0 \in B_0, \cdots, x_m \in B_m\}$ 的集合的全体, 其中 $B_i \in E, i = 0, 1, \cdots, m(m) \geqslant 0$. 显然, 这一集系 J 是 $\mathscr{F}(\mathscr{E}^\infty)$ 中集合的 $\pi-$ 系.

现在考虑第二章 §2 中定理 3 的条件.

条件 (h_1) 成立, 因为当 $A \in J$ 时, 根据上面已证明的结果有 $I_A \in \mathscr{H}$ (需要在 (5) 式中设 $H(x_0, \cdots, x_m) = I_A(x_0, \cdots, x_m)$). 由勒贝格积分的可加性, 可得条件 (h_2), 而由勒贝格积分中的单调收敛定理, 可以得到现在 (h_3).

根据上述定理 3(第二章 §2), $\mathscr{H}$ 包含一切关于 $\sigma-$ 代数 $\sigma(J)$ 的可测函数, 而根据定义 $\sigma(J)$ 就是 $\sigma-$ 代数 $\mathscr{E}^\infty = \mathscr{B}(E^\infty)$ (见第二章 §2 的第 4 和第 8 小节).

2. 强马尔可夫性的另一种推广 现在考虑马尔可夫性的第二种推广, 即将 "时间 n" 换成 "随机时间 τ " 的强马尔可夫性. (这一节开始引进的所有前提条件保持不变, 例如, $(\Omega, \mathscr{F}) = (E^\infty, \mathscr{E}^\infty), \cdots$)

以 $\tau = \tau(\omega)$ 表示有限随机变量 $\tau(\omega)$, 且对于每一个 $n \geqslant 0$, 满足

$$\{\omega : \tau(\omega) = n\} \in \mathscr{F}_n^X.$$

按照第七章 §1 引进的术语 (见定义 3), 这样的随机变量称做 (有限) 马尔可夫时间或停止时间.

将停时 τ 与 $\sigma-$ 代数流 $(\mathscr{F}_n^X)_{n \geqslant 0}$:

$$\mathscr{F}_\tau^X = \{A \in \mathscr{F}^X : A \bigcap \{\tau = n\} \in \mathscr{F}_n^X \quad 对于一切\ n \geqslant 0\},$$

相联系, 其中 $\mathscr{F}^X = \sigma(\bigcup \mathscr{F}_n^X)$ 是在 "随机区间" $[0, \tau]$ 上观测到的事件的 $\sigma-$ 代数.

定理 2 *设定理 1 中提出的全部条件成立, 而 $\tau = \tau(\omega)$ 是有限马尔可夫时间. 那么, 有如下强马尔可夫性:*

$$\mathbf{E}_\pi(H \circ \theta_\tau | \mathscr{F}_\tau^X) = \mathbf{E}_{X_\tau} H \quad (\mathbf{P}_\pi - \text{a.c.}). \tag{7}$$

在进行证明之前, 首先关于应当如何理解 $\mathbf{E}_{X_\tau} H$ 和 $H \circ \theta_\tau$, 作一些说明.

记 $\psi(x) = \mathbf{E}_x H$. (在第 1 小节已经指出, $\psi(x)$ 是 $\mathscr{E}-$ 可测函数.) 把 $\mathbf{E}_{X_\tau} H$ 理解为 $\psi(X_\tau) = \psi(X_{\tau(\omega)}(\omega))$ 的值. 关于 $(H \circ \theta_\tau)(\omega)$, 应理解为 $(H \circ \theta_{\tau(\omega)})(\omega) = H(\theta_{\tau(\omega)}(\omega))$.

证明 设集合 $A \in \mathscr{F}_\tau$. 像定理 1 一样, 为证明 (7) 式, 需要证明

$$\mathbf{E}_\pi(H \circ \theta_\tau; A) = \mathbf{E}_\pi(\mathbf{E}_{X_\tau} H; A). \tag{8}$$

等式的左侧为

$$\begin{aligned}\mathbf{E}_\pi(H\circ\theta_\tau;A)&=\sum_{n=0}^{\infty}\mathbf{E}_\pi(H\circ\theta_\tau;A\bigcap\{\tau=n\})\\&=\sum_{n=0}^{\infty}\mathbf{E}_\pi(H\circ\theta_n;A\bigcap\{\tau=n\}).\end{aligned}\tag{9}$$

由 (8) 式的右侧, 得

$$\mathbf{E}_\pi(\mathbf{E}_{X_\tau}H;A)=\sum_{n=0}^{\infty}\mathbf{E}_\pi(\mathbf{E}_{X_n}H;A\bigcap\{\tau=n\}).\tag{10}$$

事件 $A\bigcap\{\tau=n\}\in\mathscr{F}_n^X$. 由于 (4) 式, (9) 式和 (10) 式的右侧相等, 可见强马尔可夫性 (7) 得证. □

系　设函数 $H(x_0,x_1,\cdots)=I_A(x_0,x_1,\cdots)$, 而 $A=\{\omega:(x_0,x_1,\cdots)\in B\},B\in\mathscr{E}^\infty=\mathscr{B}(E^\infty)$, 则由 (7) 式得强马尔可夫性的如下常用形式:

$$\begin{aligned}&\mathbf{P}_\pi\{\omega:(X_\tau,X_{\tau+1},\cdots)\in B|X_0,X_1,\cdots,X_\tau\}\\&=\mathbf{P}_{X_\tau}\{(X_0,X_1,\cdots)\in B\}\quad(\mathbf{P}_\pi-\text{a.c.}).\end{aligned}\tag{11}$$

注 1　如果分析一下强马尔可夫性 (7) 式的证明, 就会注意到, 实际上有如下性质.

设对于任意 $n\geqslant 0$, 定义在 $\Omega=E^\infty$ 上的实函数 $H_n=H_n(\omega)$ 为 $\mathscr{F}-$ 可测的 ($\mathscr{F}=\mathscr{E}^\infty$), 并且一致有界 (即 $|H_n(\omega)|\leqslant c,n\geqslant 0,\omega\in\Omega$). 那么, 对于每一个有限马尔可夫时间 $\tau=\tau(\omega)(\tau(\omega)<\infty,\omega\in\Omega)$, 有如下形式的强马尔可夫性:

$$\mathbf{E}_\pi(H_\tau\circ\theta_\tau|\mathscr{F}_\tau^X)=\psi(\tau,X_\tau)\quad(\mathbf{P}_\pi-\text{a.c.}).\tag{12}$$

其中 $\psi(n,x)=\mathbf{E}_xH_n$, 而 $H_\tau\circ\theta_\tau=(H_\tau\circ\theta_\tau)(\omega)=H_{\tau(\omega)}(\theta_{\tau(\omega)}(\omega))$.

注 2　上面假设 $\tau=\tau(\omega)$ 是有限马尔可夫时间. 假如不是这样, 即 $\tau(\omega)\leqslant\infty,\omega\in\Omega$, 那么 (12) 式应当改为如下形式 (练习题 3):

$$\mathbf{E}_\pi(H_\tau\circ\theta_\tau|\mathscr{F}_\tau^X)=\psi(\tau,X_\tau)\quad(\{\tau<\infty\};\mathbf{P}_\pi-\text{a.c.}).\tag{13}$$

换句话说, 在此情形下, (12) 式在集合 $\{\tau<\infty\}$ 上 $\mathbf{P}_\pi-$ a.c. 成立,

3. 强马尔可夫性的例　在研究重对数定律时, 我们曾经用到一个不等式 (第四章 §4 中的引理 1; 亦见下面 (14) 式), 对于布朗运动 $B=(B_t)_{t\leqslant T}$, 等式

$$\mathbf{P}\left\{\max_{0\leqslant t\leqslant T}B_t>a\right\}=2\mathbf{P}\{|B_T|>a\}$$

是上述不等式的类似 (见 [131] 的第 III 章).

设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量序列, 并且各随机变量共同的分布 (关于 0) 对称. 记 $X_0 = x \in \mathbb{R}, X_m = X_0 + (\xi_1 + \cdots + \xi_m), m \geqslant 1$. 像前面一样, 以 $\mathbf{P}_x$ 表示序列 $X = (X_m)_{m\geqslant 0}, (X_0 = x)$ 的概率分布. (假设空间 $\Omega = \{\omega\}$, 是由坐标给定的, 其中 $\omega = (x_0, x_1, \cdots)$, 而 $X_m(\omega) = x_m$.)

由 (很容易更改的) 第四章 §4 引理 1 的 (9) 式, 可见对于任意 $a > 0$, 有

$$\mathbf{P}_0\left\{\max_{0\leqslant m\leqslant n} X_m > a\right\} \leqslant 2\mathbf{P}_0\{X_n > a\}. \tag{14}$$

引进马尔可夫时间 $\tau = \tau(\omega)$, 设

$$\tau(\omega) = \inf\{0 \leqslant m \leqslant n : X_m(\omega) > a\}. \tag{15}$$

(一般, 设 $\inf \varnothing = \infty$.) 我们现在说明, 假如允许像处理确定性变量一样, 对待这样的 (随机) 时间, 应如何利用所引进的马尔可夫时间, 或许可以给予不等式 (14) "很容易的证明". (对照第四章 §4 引理 1 的证明) 易见

$$\begin{aligned}\mathbf{P}_0\{X_n > a\} &= \mathbf{P}_0\{(X_n - X_{\tau\wedge n}) + X_{\tau\wedge n} > a\}\\ &\geqslant \mathbf{P}_0\{X_n - X_{\tau\wedge n} \geqslant 0, X_{\tau\wedge n} > a\} = \mathbf{P}_0\{X_n - X_{\tau\wedge n} \geqslant 0\}\mathbf{P}_0\{X_{\tau\wedge n} > a\}\\ &\geqslant \frac{1}{2}\mathbf{P}_0\{X_{\tau\wedge n} > a\} = \frac{1}{2}\mathbf{P}_0\{\tau \leqslant n\} = \frac{1}{2}\mathbf{P}_0\left\{\max_{0\leqslant m\leqslant n} X_m > a\right\},\end{aligned} \tag{16}$$

其中我们利用了 "'仿佛' 几乎显然的" 性质: 随机变量 $X_n - X_{\tau\wedge n}$ 和 $X_{\tau\wedge n}$ 相互独立. 当然, 对于确定性时间 τ 这是对的; 然而, 对于随机时间 τ, 一般说来这是不对的 (练习题 4). (于是, 这里的 "很容易的证明" 不能认为是适当的.)

我们现在基于强马尔可夫性 (13) 式的应用, 引进不等式 (14) 的真正 "正确的证明".

由于 $\{X_n > a\} \subseteq \{\tau \leqslant n\}$, 可见

$$\mathbf{P}_0\{X_n > a\} = \mathbf{E}_0(I_{\{X_n > a\}}; \tau \leqslant n). \tag{17}$$

引进函数 $H_m = H_m(x_0, x_1, \cdots)$, 设

$$H_m(x_0, x_1, \cdots) = \begin{cases} 1, & 若\ m \leqslant n,\ 且\ x_{n-m} > a, \\ 0, & 其他. \end{cases}$$

由相应的定义可见, 在集合 $\{\tau \leqslant n\}$ 上, 有

$$(H_\tau \circ \theta_\tau)(x_0, x_1, \cdots) = \begin{cases} 1, & 若\ x_n > a, \\ 0, & 其他, \end{cases} \tag{18}$$

因此, 注意到 (17) 式, 由于 $\{X_n > a\} \subseteq \{\tau \leqslant n\}$, 以及 $\{\tau \leqslant n\} \in \mathscr{F}_\tau$, 可得

$$\mathbf{P}_0\{X_n > a\} = \mathbf{E}_0(H_\tau \circ \theta_\tau; \tau \leqslant n) = \mathbf{E}_0[\mathbf{E}_0(H_\tau \circ \theta_\tau | \mathscr{F}_\tau^X); \tau \leqslant n]. \tag{19}$$

由强马尔可夫性 (13), 在集合 $\{\tau \leqslant \infty\}$ 上, 有

$$\mathbf{E}_0(H_\tau \circ \theta_\tau | \mathscr{F}_\tau) = \psi(\tau, X_\tau) \quad (\mathbf{P}_0 - \text{a.c.}). \tag{20}$$

根据定义 $\psi(m,x) = \mathbf{E}_x H_m$, 而对于 $x \geqslant a$,

$$\mathbf{E}_x H_m = \mathbf{P}_x\{X_{n-m} > a\} \geqslant \mathbf{P}_x\{X_{n-m} > x\} \geqslant \frac{1}{2}$$

(由于随机变量 $\xi_1, \xi_2, \cdots$ 的分布的对称性, 得最后的不等式).

于是, 在集合 $\{\tau \leqslant n\}$ 上

$$\mathbf{E}_0(H_\tau \circ \theta_\tau | \mathscr{F}_\tau) \geqslant \frac{1}{2} \quad (\mathbf{P}_0 - \text{a.c.}). \tag{21}$$

由此以及由 (19), (20) 式得所要求证明的不等式 (14).

4. 柯尔莫戈洛夫 – 查普曼方程 如果考虑柯尔莫戈洛夫 – 查普曼方程 (13) 和方程 (38) (第一章 §12), 则可以注意到它们非常相似. 自然应分析它们的表述与它们的结论中的共同点和差异. (我们仅限于讨论具有离散状态集合 E 的齐次马尔可夫链.)

对于 $n \geqslant 1, 1 \leqslant k \leqslant n, i, j \in E$, (注意到 (1) 式和 (2) 式), 有

$$\begin{aligned}
\mathbf{P}_i\{X_n = j\} &= \sum_{\alpha \in E} \mathbf{P}_i\{X_n = j, X_k = \alpha\} + \sum_{\alpha \in E} \mathbf{E}_i I(X_n = j) I(X_k = \alpha) \\
&= \sum_{\alpha \in E} \mathbf{E}_i\{\mathbf{E}_i[I(X_n = j) I(X_k = \alpha) | \mathscr{F}_k]\} = \sum_{\alpha \in E} \mathbf{E}_i\{I(X_k = \alpha) \mathbf{E}_i[I(X_{n-k} = j) | \mathscr{F}_k]\} \\
&\overset{(1)}{=} \sum_{\alpha \in E} \mathbf{E}_i\{I(X_k = \alpha) \mathbf{E}_i[I(X_{n-k} = j) \circ \theta_k | \mathscr{F}_k]\} \\
&\overset{(2)}{=} \sum_{\alpha \in E} \mathbf{E}_i\{I(X_k = \alpha) \mathbf{E}_{X_k} I(X_{n-k} = j)\} \\
&= \sum_{\alpha \in E} \mathbf{E}_i\{I(X_k = \alpha) \mathbf{E}_\alpha I(X_{n-k} = j)\} = \sum_{\alpha \in E} \mathbf{E}_i I(X_k = \alpha) \mathbf{E}_\alpha I(X_{n-k} = j) \\
&= \sum_{\alpha \in E} \mathbf{P}_i(X_k = \alpha) \mathbf{P}_\alpha(X_{n-k} = j),
\end{aligned} \tag{22}$$

这就是柯尔莫戈洛夫 – 查普曼方程 (13), 而在第一章 §12 中表示为

$$p_{ij}^{(n)} = \sum_{\alpha \in E} p_{i\alpha}^{(k)} p_{\alpha j}^{(n-k)}.$$

假如在 (22) 式中, 将时间 k 换成 (取 $1, 2, \cdots, n$ 为值的) 马尔可夫时间 τ, 并且将采用马尔可夫性 (2) 换成使用强马尔可夫性 (7), 则得 (练习题 5) 柯尔莫戈洛夫 – 查普曼方程如下自然的形式:

$$\mathbf{P}_i\{X_n = j\} = \sum_{\alpha \in E} \mathbf{P}_i\{X_\tau = \alpha\} \mathbf{P}_\alpha\{X_{n-\tau} = j\}. \tag{23}$$

在 (22) 式和在 (23) 式中都是对相变量 $\alpha \in E$ 求和. 在第一章 §12 的 (38) 式中是对时间变量求和.

注意到上述情况, 现在假设 τ 是取 $1, 2, \cdots, n$ 为值的马尔可夫时间. 像上述 (38) 式一样, 我们开始推导下面的式子, 有

$$\begin{aligned}
\mathbf{P}_i\{X_n = j\} &= \sum_{k=1}^{n} \mathbf{P}_i\{X_n = j, \tau = k\} + \mathbf{P}_i\{X_n = j, \tau \geqslant n+1\} \\
&= \sum_{k=1}^{n} \mathbf{E}_i I(X_n = j) I(\tau = k) + \mathbf{P}_i\{X_n = j, \tau \geqslant n+1\} \\
&= \sum_{k=1}^{n} \mathbf{E}_i\{\mathbf{E}_i[I(X_n = j) I(\tau = k) | \mathscr{F}_k]\} + \mathbf{P}_i\{X_n = j, \tau \geqslant n+1\} \\
&= \sum_{k=1}^{n} \mathbf{E}_i\{I(\tau = k) \mathbf{E}_i I[(X_n = j) | \mathscr{F}_k]\} + \mathbf{P}_i\{X_n = j, \tau \geqslant n+1\} \\
&= \sum_{k=1}^{n} \mathbf{E}_i\{I(\tau = k) \mathbf{E}_i[I(X_{n-k} = j) \circ \theta_k | \mathscr{F}_k]\} + \mathbf{P}_i\{X_n = j, \tau \geqslant n+1\} \\
&= \sum_{k=1}^{n} \mathbf{E}_i\{I(\tau = k) \mathbf{E}_i I(X_{n-k} = j)\} + \mathbf{P}_i\{X_n = j, \tau \geqslant n+1\}. \qquad (24)
\end{aligned}$$

在第一章 §12 的第 7 小节, 如果集合 $\{\cdot\} = \varnothing$, 则时间 $\tau = \tau_j$, 其中

$$\tau_j = \min\{1 \leqslant k \leqslant n : X_k = j\}$$

满足条件 $\tau_j = n+1$. 从而, 在这种情形下, (24) 式自然可以简化为:

$$\begin{aligned}
\mathbf{P}_i\{X_n = j\} &= \sum_{k=1}^{n} \mathbf{E}_i[I(\tau_j = k) \mathbf{E}_{X_{\tau_j}} I(X_{n-k} = j)] \\
&= \sum_{k=1}^{n} \mathbf{E}_i[I(\tau_j = k) \mathbf{E}_j I(X_{n-k} = j)] \\
&= \sum_{k=1}^{n} \mathbf{E}_i I(\tau_j = k) \mathbf{E}_j I(X_{n-k} = j) \\
&= \sum_{k=1}^{n} \mathbf{P}_i\{\tau_j = k\} \mathbf{P}_j\{X_{n-k} = j\},
\end{aligned}$$

成为第一章 §12 的 (38) 式

$$p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)}. \qquad (25)$$

在 (24) 式中也可以得到另一有用的公式, 其中 (与柯尔莫戈洛夫 – 查普曼方程不同) 是对时间变量求和. 例如, 假设马尔可夫时间

$$\tau(\alpha) = \min\{1 \leqslant k \leqslant n : X_k = \alpha(k)\},$$

以及 (确定性) 函数 $\alpha = \alpha(k), 1 \leqslant k \leqslant n$, 而马尔可夫链满足条件: 对于给定的 i 和 $n, \mathbf{P}_i\{\tau(\alpha) \leqslant n\} = 1$. 那么, 由 (24) 式可见

$$\mathbf{P}_i\{X_n = j\} = \sum_{k=1}^{n} \mathbf{E}_i[I(\tau(\alpha) = k)\mathbf{E}_{X_{\tau(\alpha)}} I(X_{n-k} = j)]$$
$$= \sum_{k=1}^{n} \mathbf{E}_i I(\tau(\alpha) = k)\mathbf{E}_{\alpha(k)} I(X_{n-k} = j),$$

即 (对照 (23) 式)

$$\mathbf{P}_i\{X_n = j\} = \sum_{k=1}^{n} \mathbf{P}_i\{\tau(\alpha) = k\}\mathbf{P}_{\alpha(k)}\{X_{n-k} = j\}.$$

5. 练习题

1. 证明第 1 小节的注中的函数 $\psi(x) = \mathbf{E}_x H$ 为 $\mathscr{E}-$ 可测.
2. 证明性质 (12).
3. 证明性质 (13).
4. 问第 3 小节的例子中, 随机变量 $X_n - X_{\tau\wedge n}$ 和 $X_{\tau\wedge n}$ 是否独立?
5. 证明性质 (23).

§3. 马尔可夫链的极限、遍历和平稳概率分布问题

1. 广义马尔可夫性 我们在 §1 中已经指出, 用马尔可夫链描绘的*无后效*随机系统的*渐近性质*问题, 是马尔可夫过程理论的中心课题之一. 这在一定程度上与下面的情况有关, 在相当广泛的条件下, 这样的系统好像趋向于 "*稳定*", 进入平稳 "*状态*".

可以从不同的角度研究齐次马尔可夫链 $X = (X_n)_{n\geqslant 0}$ 的*极限性质*. 例如, 可以像狭义平稳序列的遍历性定理那样 (第五章 §3 定理 3), 对不同函数 $f = f(x)$, 研究当 $n \to \infty$ 时形如 $n^{-1}\sum\limits_{m=0}^{n-1} f(X_m)$ 泛函 $\mathbf{P}_\pi-$ 几乎必然收敛问题. 像第一章那样, 大数定律成立的条件也很重要.

进一步的叙述, 主要的注意力, 并不放在诸如*几乎必然收敛*或*依概率收敛*上, 而放在当 $n \to \infty$ 时 n *步转移概率* $P^{(n)}(x; A)$ *的渐近性质*的问题上 (见 §1 的 10 式), 以及*平稳 (不变) 测度* $q = q(A)$ 存在性的问题上, 其中测度

$$q(A) = \int P(x; A)q(\mathrm{d}x), \tag{1}$$

其中 $P(x; A)$ 是 (一步) 转移函数.

需要强调, 在 (1) 式定义中, 一般将完全*没有*假定测度 $q = q(A)$ 是概率测度 (没有要求 $q(E) = 1$).

假如 $q = q(A)$ 是概率测度, 则习惯上称之为平稳分布或不变分布. 这一术语的含义是完全清楚的: 如果把分布 q 取为初始分布 π, 即认为 $\mathbf{P}_q\{X_0 \in A\} = q(A)$, 则由 (1) 式知, 对于任意 $n \geqslant 1, \mathbf{P}_q\{X_n \in A\} = q(A)$, 即该分布随时间是不变的.

不难举出不存在平稳分布 $q = q(A)$, 但是存在平稳测度的例子.

例 设 $X = (X_n)_{n\geqslant 0}$ 是由伯努利概型生成的马尔可夫链, 即 $X_{n+1} = X_n + \xi_{n+1}$, 其中 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量序列: $\mathbf{P}\{\xi_n = +1\} = p, \mathbf{P}\{\xi_n = -1\} = q$. 设 $X_0 = x$, 其中 $x \in \{0, \pm 1, \cdots\}$. 显然这时转移函数为

$$P(x; \{x+1\}) = p, \quad P(x; \{x-1\}) = q.$$

不难验证, 对于任意 $x \in \{0, \pm 1, \cdots\}$, 满足 $q(\{x\}) = 1$ 测度 $q(A)$, 是 (1) 式的解之一. 如果 $p \neq q$, 则满足 $q(\{x\}) = (p/q)^x$ 测度 $q(A)$ 是第二不变测度. 显然, 这些测度都非概率测度, 而且这里不存在概率不变测度.

这一简单的例子说明, 为存在平稳 (不变) 分布, 需要所研究的马尔可夫链满足一定条件. 当 $n \to \infty$ 时, 关于转移概率 $P^{(n)}(x; A)$ 的极限值的问题之所以重要, 首先是因为不依赖于初始状态 x 的极限的存在问题. 这时应该注意到完全可能出现任何极限分布都不存在的情形. 例如, 可能出现这样的情形: 对于任何 $A \in \mathscr{E}$ 和任意初始状态 $x \in E, \lim P^{(n)}(x; A) = 0$. 例如, 在上面的例子中只需设 $p = 1$, 即考虑确定性的向右运动 (亦见 §8 中的例 4 和例 5; 对照 §5 的练习题 6).

对于任意相空间 $(E, \mathscr{E})$ 的情形, 寻找平稳 (不变) 分布, 以及寻找 (具有某种性质的) 转移概率极限值的分布, 是相当困难的课题 (例如, 见 [104]). 然而, 对于可数状态空间 ("可数马尔可夫链") 的情形, 这里得到了重要的和相当明显表达的结果. 有关结果将在 §6 和 §7 中介绍. 不过, 我们需要首先对可数状态马尔可夫链, 按转移概率的代数性质和渐近性质, 进行详细的分类.

需要指出, 所考虑的关于平稳分布和极限 $\lim P^{(n)}(x; A)$ 的存在性问题, 二者密切相关. 事实上, 如果 $\lim P^{(n)}(x; A) (= v(A))$ 存在, 与 x 无关, 而且 (对于 $A \in \mathscr{E}$) 是测度, 则由柯尔莫戈洛夫 – 查普曼方程

$$P^{(n+1)}(x; A) = \int P^{(n)}(x; \mathrm{d}y) P(y; A)$$

当 $n \to \infty$ 时, 经 (形式地) 极限过程, 得

$$v(A) = \int P(y; A) v(\mathrm{d}y).$$

于是, $v = v(A)$ 是平稳 (不变) 测度.

2. 要研究的主要问题 以下到处假设, 所考虑的马尔可夫链 $X = (X_n)_{n\geqslant 0}$, 在可数相空间 $E = \{1, 2, \cdots\}$ 取值. 为简便计, 我们将用 $p_{ij} (i, j \in E)$ 表示转移函数 $P(i, \{j\})$. (为直观计, 求游动 "质点") 从状态 i 到状态 j 的转移概率记为 $p_{ij}^{(n)}$.

我们关心的主要问题涉及如下条件的说明:

A. 对于一切 $j \in E$, 存在不依赖于初始状态 $i \in E$ 的极限

$$\pi_j = \lim_n p_{ij}^{(n)};$$

B. 这些 $\mathbb{\Pi} = (\pi_1, \pi_2, \cdots)$ 的极限值是概率分布, 即 $\pi_j \geqslant 0, \sum_{j \in E} \pi_j = 1$;

C. 链 $X = (X_n)_{n \geqslant 0}$ 是遍历的, 换句话说, 极限值 $\mathbb{\Pi} = (\pi_1, \pi_2, \cdots)$ 满足条件一切 $\pi_j > 0, \sum_{j \in E} \pi_j = 1$;

D. 平稳 (不变) 概率分布 $\mathbb{Q} = (q_1, q_2, \cdots)$ 存在并且唯一, 即 $\mathbb{Q} = (q_1, q_2, \cdots)$ 满足条件 $q_j \geqslant 0, \sum_{j \in E} q_j = 1$, 并且对于一切 $j \in E$, 有

$$q_j = \sum_{i \in E} q_i p_{ij}.$$

注 这里用到的术语 "遍历性" 在第五章已经出现过 (遍历性作为度量的可递性, 辛钦 – 博赫纳遍历性定理). 从字面上讲, 这些词的含义涉及不同的对象, 然而它们也有共同点, 就是它们都反映, 在时间参数趋向无穷的情况下, 各种概率特征的极限性质.

3. 练习题

1. 列举马尔可夫链的例子, 使得 $\pi_j = \lim_n p_{ij}^{(n)}$ 存在, 并且: (a) 不依赖于初始状态 j; (b) 依赖于初始状态 j.

2. 列举遍历链和非遍历链的例子.

3. 举出不是遍历分布的平稳分布例子.

§4. 马尔可夫链的状态按转移概率矩阵的代数性质分类

1. 转移概率矩阵 假设所考虑的马尔可夫链有可数个状态集 $E = \{1, 2, \cdots\}$, 而其转移概率为 $p_{ij}, i, j \in E$. 以 $\mathbb{P} = (p_{ij})$ 表示这些转移概率构成的矩阵 (表), 或者用展开的形式将 $\mathbb{P}$ 写成

$$\mathbb{P} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ \vdots & \vdots & \vdots & \\ p_{i1} & p_{i2} & p_{i3} & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}.$$

下面要引进的马尔可夫链状态的分类, 完全决定于转移概率矩阵 $\mathbb{P}$ 及其 $n \geqslant 1$ 次幂 $\mathbb{P}^{(n)}$ 的代数性质.

转移概率矩阵 $\mathbb{P}$ 完全决定于状态到状态的一步转移. 而由于马尔可夫性, 矩阵 $\mathbb{P}^{(n)} = (p_{ij}^{(n)})$ 决定 n 步的转移. 例如, 矩阵

$$\mathbb{P} = \begin{pmatrix} 1/2 & 1/2 \\ 0 & 0 \end{pmatrix}$$

和与其相对应的图形 (第一章 §12) 表明, 它所决定的 "质点" 的运动是这样的: "质点" 在状态 0 和 1 上游动: 经一步以概率可能 $0 \to 1$ 移动 (以概率 1/2), 但是移动 $1 \to 0$ 不可能. 由 n 步转移概率矩阵

$$\mathbb{P}^{(n)} = \begin{pmatrix} 2^{-n} & 1-2^{-n} \\ 0 & 1 \end{pmatrix}$$

可见对于任意 $n \geqslant 1, p_{10}^{(n)} = 0$. 因此, 移动 $1 \to 0$ 显然不可能, 可见无论经过任何步移动, $1 \to 0$ 当然也不可能.

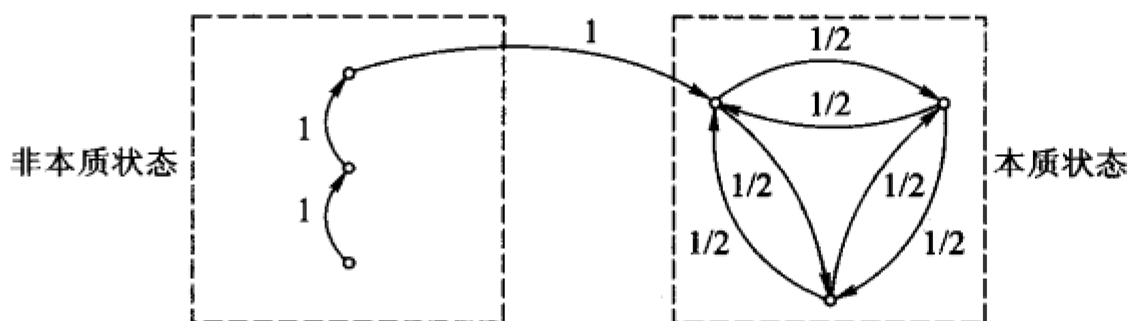


图 36

在该例中从状态 0 进入状态 1, 但是不可能再越出状态 1.

由图 36, 可见根据该图很容易复原转移概率矩阵 $\mathbb{P}$. 由此图的形态可以看出, 这里有三个状态 (图的左半部分), 从某一状态越出后, 就再也不能返回.

从按该图形游动 "质点" 的 "将来" 的性质看, 这三个状态*不重要* (故称之为*非本质状态*), 因为, 从这样的状态*可以越出*, 但是*不可能再返回*.

这样的 "非本质" 状态不值得研究, 立即可以去掉, 并把全部注意力放在剩下的 "本质" 状态的分类上. (可以用转移概率性质的语言: $p_{ij}^{(n)}, i, j \in E, n \geqslant 1$, 给 "本质" 状态和 "非本质" 状态的描述性定义以确切的形式, 练习题 1.)

2. 可达状态和可通状态 为将本质状态或这样的状态组分类, 需要用到如下概念.

定义 1 称状态 j 为由状态 i *可达的*, 记作 $i \to j$, 如果存在 $n \geqslant 0$, 使 $p_{ij}^{(n)} > 0$ (若 $i = j$, 则 $p_{ij}^{(0)} = 1$; 若 $i \neq j$, 则 $p_{ij}^{(0)} = 0$).

称状态 i 和 j 为*互通的*, 记作 $i \leftrightarrow j$, 如果 $i \to j$ 且 $j \to i$, 即状态 i 和 j *互为可达的*.

引理 1 *互通性 $i \leftrightarrow j$ (相互可达性) 是 (转移矩概率阵为 $\mathbb{P}$ 的马尔可夫链的) 状态的等价关系.*

证明　根据等价关系 $i \leftrightarrow j$ 的定义, 需要验证其**自反性** ($i \leftrightarrow i$), **对称性** (即若 $i \leftrightarrow j$, 则 $j \leftrightarrow i$) 和**可递性** (即若 $i \leftrightarrow j, j \leftrightarrow k$, 则 $i \leftrightarrow k$).

前两条性质可以由状态的**互通性**得到. 关于可递性由可尔莫戈洛夫 -- 查普曼方程, 当 $p_{ij}^{(n)} > 0, p_{jk}^{(m)} > 0$ 时, 由

$$p_{jk}^{(n+m)} = \sum_{l \in E} p_{il}^{(n)} p_{lk}^{(m)} \geqslant p_{ij}^{(n)} p_{jk}^{(m)} > 0,$$

可见 $i \to k$. 类似地 $k \to i$. 于是 $i \leftrightarrow k$. □

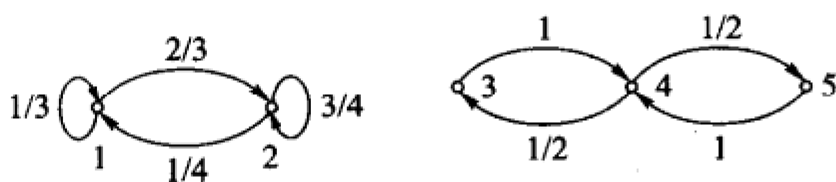
我们把所有互通状态 $i, j, k, \cdots (i \leftrightarrow j, j \leftrightarrow k, k \leftrightarrow i, \cdots)$ 都归为一类. 那么, 所有这样的状态类或者重合, 或者不相交. 从而, **互通性**关系把 (本质) 状态的整个集合 E, 分割为有限或可数个不相交集合 $E_1, E_2, \cdots (E = E_1 + E_2 + \cdots)$.

我们把这些集合 $E_1, E_2, \cdots$ 称为 (本质连通) 状态的**不可约类**. 所有状态形成一个不可约类的马尔可夫链称做**不可约的**.

为演示上面引进的概念, 我们考虑马尔可夫链: 其状态空间为 $E = \{1, 2, 3, 4, 5\}$, 而转移概率矩阵为

$$\mathbb{P}^{(n)} = \left(\begin{array}{cc|ccc} 1/3 & 2/3 & 0 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right) = \left(\begin{array}{c|c} \mathbb{P}_1 & 0 \\ \hline 0 & \mathbb{P}_2 \end{array}\right).$$

这一链具有 5 个状态的图有如下形状:



显然, 该马尔可夫链有**两个不可约的状态类**: $E_1 = \{1, 2\}$ 和 $E_2 = \{3, 4, 5\}$. 研究该链的性质归结为, 研究两个链中每一个链的性质: 两个链的状态空间相应为 E_1 和 E_2, 而转移概率矩阵分别等于 $\mathbb{P}_1$ 和 $\mathbb{P}_2$.

现在考虑任何一个不可约状态类 E. 作为例子, 假设将状态类描述在图 37 中. 注意, 这里只有经过**偶数步**返回每一个状态才是可能的, 经**奇数步**只能转移到相邻状态, 而转移概率矩阵具有分块的结构:

$$\mathbb{P} = \left(\begin{array}{cc|cc} 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \\ \hline 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \end{array}\right)$$

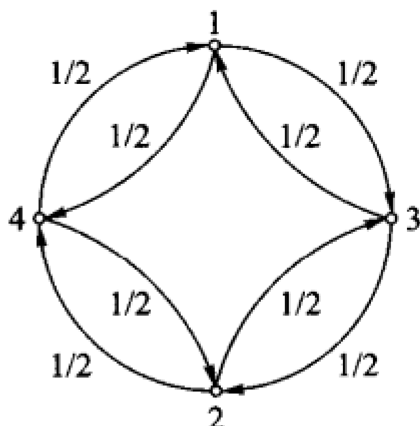


图 37 周期 $d=2$ 的马尔可夫链

由此可见状态类 $E=\{1,2,3,4\}$ 划分为两个子类 $C_0=\{1,2\}$ 和 $C_1=\{3,4\}$, 且具有如下周期性: 质点一步从子类 C_0 越出, 一定转移到子类 C_1, 而由 C_1 转移到 C_0.

3. 按周期对状态分类 上面的例子表明, 看来在一般情形下, 也可以进行不可约状态类按周期子类的分类.

为此需要引进某些概念, 以及数论中的一个结果.

定义 2 设 $\varphi=(\varphi_1,\varphi_2,\cdots)$ 是某一非负数 $\varphi_n \geqslant 0(n \geqslant 1)$ 的序列; M_φ 是使 $\varphi_n>0(n \geqslant 1)$ 的下标 n 的集合, 并且当 $\varphi_n=0(n \geqslant 1)$ 时, 设 $M_\varphi=\varnothing$, 而设 $\mathrm{GCD}(M_\varphi)=0$, 其中 $\mathrm{GCD}(M_\varphi)$ 是集合 M_φ 的最大公约数[①]. 那么, 称数

$$d(\varphi)=\mathrm{GCD}\{n \geqslant 1: \varphi_n>0\}$$

为序列 φ 的周期.

换句话说, 称序列 φ 的周期为 $d(\varphi)$, 如果由 $\varphi_n>0$ 可见 n 被 $d(\varphi)$ 整除 (即对于某个 $k \geqslant 1$, n 应具有 $d(\varphi)k$ 的形式), 则 $d(\varphi)$ 在一切具有该性质 (即对于某个整数 $l \geqslant 1, n=dl$) 的数中是最大的.

例如, 假如序列 $\varphi=(\varphi_1,\varphi_2,\cdots)$ 满足:

$$\varphi_n=\begin{cases}\varphi_{4k}>0, & \text{若 } n=4k, k=1,2,\cdots, \\ 0, & \text{若 } n \neq 4k,\end{cases}$$

则序列 φ 的周期为 $d(\varphi)=4$; 虽然 $\varphi_{2l}>0(l=2,4,8)$, 但是 $d(\varphi) \neq 2$.

定义 3 称序列 $\varphi=(\varphi_1,\varphi_2,\cdots)$ 为非周期的, 如果其周期 $d(\varphi)=1$.

下面在根据周期性对状态分类时, 将要用到数论中的如下初等结果.

引理 2 设 M 是关于加法运算封闭的非负整数的集合 $(M \subseteq E)$, 并且 $\mathrm{GCD}(M)=1$. 那么, 对于某一 n_0, 所有数 $n \geqslant n_0$ 都属于集合 M.

把序列 $\varphi=(\varphi_1,\varphi_2,\cdots)$, 取为序列 $(p_{jj}^{(1)},p_{jj}^{(2)},\cdots)$ 或 $(p_{jj}^{(d)},p_{jj}^{(2d)},\cdots), d \geqslant 1$, 其中 j 是马尔可夫链的某个状态, 而马尔可夫链转移概率矩阵为 $\mathbb{P}=(p_{ij})$, 且 $p_{jj}^{(n)}$ 是矩阵 $\mathbb{P}^{(n)}(n \geqslant 1)$ 的元素, $\mathbb{P}^{(1)}=\mathbb{P}$. (这时, 如果 $d(j)$ 是序列 $(p_{jj}^{(1)},p_{jj}^{(2)},\cdots)$ 的周期, 则称状态 j 有周期 $d(j)$.) 那么, 就有如下结果.

①GCD是 "Greatest Common Divisor" (最大公约数) 的缩写. —— 译者

定理 1 设状态 j 的周期 $d = d(j)$.

如果 $d = 1$, 则存在 $n_0 = n_0(j, d)$, 使对所有 $n \geqslant n_0$ 转移概率 $p_{jj}^{(n)} > 0$.

如果 $d > 1$, 则存在 $n_0 = n_0(j, d)$, 使对所有 $n \geqslant n_0$ 转移概率 $p_{jj}^{(nd)} > 0$.

如果 $d \geqslant 1$ 且对于某个 $i \in E$ 和 $m > 1, p_{ii}^{(m)} > 0$, 则存在 $n_0 = n_0(j, d, m)$, 使对所有 $n \geqslant n_0$ 转移概率 $p_{ij}^{(m+nd)} > 0$.

现在给出一定理, 它说明不可约类的状态的周期, 具有 "单一类型" 的特性.

定理 2 设 $E_* = \{i, j, \cdots\}$ 是集合 E 中 (互通) 状态的某一不可约类.

这样状态类的一切状态按如下意义是 "单一类型的": 它们有同一周期 (记作 $d(E_*)$, 并称为 E_* 类的周期.)

证明 设 $i, j \in E_*$, 则存在 k 和 l, 使 $p_{ij}^{(k)} > 0$ 和 $p_{ji}^{(l)} > 0$. 而根据柯尔莫戈洛夫 - 查普曼方程

$$p_{ii}^{(k+l)} = \sum_{\alpha \in E} p_{i\alpha}^{(k)} p_{\alpha i}^{(l)} \geqslant p_{ij}^{(k)} p_{ji}^{(l)} > 0,$$

因此 $k + l$ 应被状态 $i \in E_*$ 的周期 $d(i)$ 整除. 设 $d(j)$ 是状态 $j \in E_*$ 的周期, 由于 $d(j) = \mathrm{GCD}\{n : p_{jj}^{(n)} > 0\}$, 可见 n 应该被 $d(j)$ 整除, 而因为

$$p_{ii}^{(n+k+l)} \geqslant p_{ij}^{(k)} p_{jj}^{(n)} p_{ji}^{(l)} > 0,$$

故 $n + k + l$ 被 $d(i)$ 整除; 由于 $k + l$ 被整除, 而且 $d(j) = \mathrm{GCD}\{n : p_{ij}^{(n)} > 0\}$, 因而 $d(i) \leqslant d(j)$.

由对称性有 $d(j) \leqslant d(i)$, 于是 $d(i) = d(j)$. □

4. 不可约马尔可夫链 假如状态集合组成 (互通状态的) 不可约类 $E_* \subseteq E$, 且 $d(E_*) = 1$, 则这样的类常称做非周期状态类.

现在考虑 $d(E_*) > 1$ 的情形.

在这样的类内, 由状态到状态的转移, 可以以相当出奇别致的方式实现 (就像上面讨论过的、周期为 $d(E_*) = 2$ 的马尔可夫链的例子一样; 见图 37). 然而, 结果表明在由一组状态向另一组状态的转移中, 具有完全的 "周期性 (循环性)".

定理 3 设 $E_*(E_* \subseteq E)$ 是周期为 $d = d(E_*) > 1$ 的不可约状态类.

那么, 存在称为循环子类的 d 个状态组 $C_0, C_1, \cdots, C_{d-1}(E_* = C_0 + C_1 + \cdots + C_{d-1})$, 其特征是: 在时间 $n = p + kd(p = 0, 1, \cdots, d-1; k = 0, 1, \cdots)$, "质点" 将处于子类 C_p 中, 并且在下一时间进入 C_{p+1}, 然后进入 $C_{p+2}, \cdots$, 进入 C_{d-1}, 由 C_{d-1} 进入 C_0 等.

证明 固定某个状态 $i_0 \in E_*$, 并引进如下子类:

$$C_0 = \{j \in E_* : \text{若 } p_{i_0 j}^{(n)} > 0, \ \text{则 } n = kd, \ \ k = 0, 1, \cdots\},$$

$$C_1 = \{j \in E_* : \text{若 } p_{i_0 j}^{(n)} > 0, \ \text{则 } n = kd + 1, k = 0, 1, \cdots\},$$

$$\cdots\cdots\cdots$$

$$C_{d-1} = \{j \in E_* : \text{若 } p_{i_0 j}^{(n)} > 0, \ \text{则 } n = kd + (d-1), \ \ k = 0, 1, \cdots\}.$$

显然, $E_* = C_0 + C_1 + \cdots + C_{d-1}$. 现在证明, "质点" 从子类到子类的运动是按定理 3 所描述的方式进行的; 见图 38.

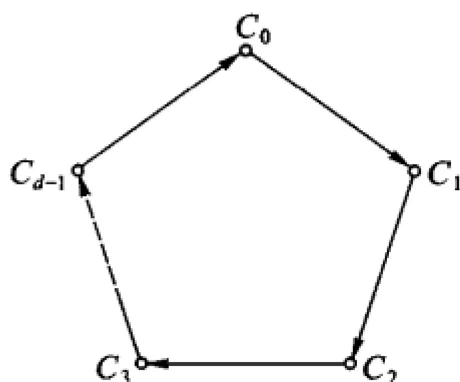
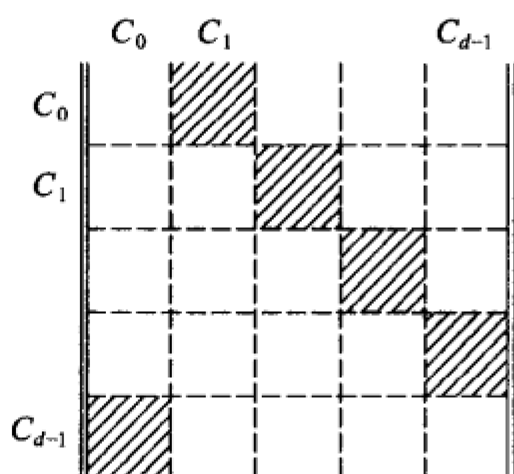


图 38 在循环子类中的运动

事实上, 我们考虑某个状态 $i \in C_p$, 并假设对于状态 $j \in E_*, p_{ij} > 0$, 则证明必有 $j \in C_{(p+1)(\mathrm{mod})}$.

假设 n 使 $p_{i_0 j}^{(n)} > 0$. 那么, n 可以表示为 $n = p + kd, p = 0, 1, \cdots, d-1$ 和 $k = 0, 1, \cdots$, 故 $n \equiv p$ (mod), 所以 $n + 1 \equiv p + 1$ (mod). 因此 $p_{i_0 j}^{(n+1)} > 0$ (根据周期的定义 $d = d(E_*)$), 故 $j \in C_{(p+1)(\mathrm{mod})}$, 而这正是要求证明的. □

需要指出, 由以上的讨论可见, 转移概率矩阵 $\mathbb{P}$ 具有分块结构:



现在假设游动 "质点" 的运动由矩阵 $\mathbb{P}$ 控制, "质点" 自子类 C_0 的某种状态开始运动. 那么, (由于子类 $C_0, C_1, \cdots, C_{d-1}$ 的定义) 此 "质点" 在 $n = p + kd$ 中的每一时间处于集合 C_p 中.

从而, 与每一个这样的状态集合 C_p, 都可以与一新的马尔可夫链相联系, 其转移概率矩阵为 $(p_{ij}^{(d)}), i, j \in C_p$. 这一新马尔可夫链是不可约的和非周期的.

于是, 由上述 (非本质状态和本质状态, 以及不可约类和循环子类) 的分类 (见综合图 39), 可以得出这样的结论:

设 $p_{ij}^{(n)}, n \geqslant 1, i, j \in E$, 是决定 "马尔可夫质点" 游动的转移概率. 在研究上述转移概率的极限性质时, 可以仅局限于 "相空间 E 本身是唯一不可约非周期状态类" 情形的研究.

在此假设下, 具有这样相空间和转移概率矩阵 $\mathbb{P}$ 的马尔可夫链 $X=(X_n)_{n\geqslant 0}$ 称做不可约的和非周期的.

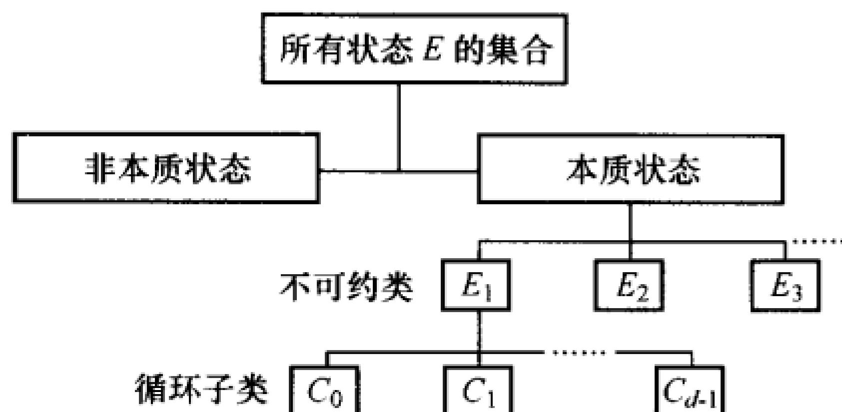


图 39　马尔可夫链的状态按转移概率 $p_{ij}^{(n)}$ 的渐近性分类

5. 练习题

1. 在第 1 小节的末尾, 讨论了非本质状态和本质状态的描述性定义. 试用转移概率 $p_{ij}^{(n)}, i,j\in E, n\geqslant 1$ 的术语, 给出非本质状态和本质状态定义的确切表述.

2. 设 $\mathbb{P}$ 是不可约的有限马尔可夫链的转移概率矩阵. 且 $\mathbb{P}^2=\mathbb{P}$. 讨论矩阵 $\mathbb{P}$ 的构造.

3. 设 $\mathbb{P}$ 是有限状态马尔可夫链 $X=(X_n)_{n\geqslant 0}$ 的转移概率矩阵; $\sigma_1,\sigma_2,\cdots$ 是独立同分布非负整数随机变量序列, 并且与 X 独立; $\tau_0=0, \tau_n=\sigma_1+\cdots+\sigma_n, n\geqslant 1$.

(a) 证明: 序列 $\widetilde{X}=(\widetilde{X}_n)_{n\geqslant 0}$ 是马尔可夫链, 其中 $\widetilde{X}_n=X_{\tau_n}$;

(b) 求马尔可夫链 $\widetilde{X}=(\widetilde{X}_n)_{n\geqslant 0}$ 的转移概率矩阵 $\widetilde{\mathbb{P}}$;

(c) 证明, 如果对于马尔可夫链 X, 状态 i 和 j 互通, 则对于马尔可夫链 $\widetilde{X}$, 状态 i 和 j 也互通.

4. 考虑马尔可夫链: 其状态空间为 $E=\{0,1\}$, 而转移概率矩阵为

$$\mathbb{P}=\begin{pmatrix}\alpha & 1-\alpha\\ 1-\beta & \beta\end{pmatrix},\quad 0<\alpha<1,\quad 0<\beta<1.$$

试描绘矩阵 $\mathbb{P}^{(n)}, n\geqslant 2$ 的构造.

§5. 马尔可夫链的状态按转移概率矩阵的渐近性质分类

1. 常返和非常返状态的概念与准则　设 $X=(X_n)_{n\geqslant 0}$ 是有可数个状态集合 $E=\{1,2,\cdots\}$ 的齐次马尔可夫链, 其转移概率为 $p_{ij}=\mathbf{P}_i\{X_1=j\}, i,j\in E$.

设

$$f_{ii}^{(n)}=\mathbf{P}_i\{X_n=i, X_k\neq i, 1\leqslant k\leqslant n-1\},\tag{1}$$

而对于 $i\neq j$,

$$f_{ij}^{(n)}=\mathbf{P}_i\{X_n=j, X_k\neq j, 1\leqslant k\leqslant n-1\}.\tag{2}$$

显然, $f_{ii}^{(n)}$ 是 "质点" 恰好在第 n 步首返状态 i 的概率, 而 $f_{ij}^{(n)}$ 是 "质点" 在 $X_0 = i$ 的条件下, 恰好在第 n 步首达状态 j 的概率.

假如设

$$\sigma_i(\omega) = \inf\{n \geqslant 1 : X_n(\omega) = i\}, \tag{3}$$

其中当上式右侧大括号 $\{\cdot\} = \varnothing$ 时 $\sigma_i(\infty) = \infty$, 则概率 $f_{ii}^{(n)}$ 和 $f_{ij}^{(n)}$ 亦可表示为下面的形式:

$$f_{ii}^{(n)} = \mathbf{P}_i\{\sigma_i = n\}, \quad f_{ij}^{(n)} = \mathbf{P}_i\{\sigma_j = n\}. \tag{4}$$

对于 $i, j \in E$, 引进变量

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}. \tag{5}$$

由 (4) 式可见

$$f_{ij} = \mathbf{P}_i\{\sigma_j < \infty\}. \tag{6}$$

换句话说, f_{ij} 是 "质点" 自状态 i 出发开始游动, 迟早到达状态 j 的概率.

在以后特别重要的是, "质点" 自状态 i 出发开始游动, 迟早返回状态 i 的概率 f_{ii}. 下面给出这些概率的定义.

定义 1 如果 $f_{ii} = 1$, 则称状态i 为*常返的* (循环的, 返回的 [recurrent], 回归的 [persistent]).

定义 2 如果 $f_{ii} < 1$, 则称状态 i 为*非常返的* (非可迁的, 非返回的, 非回归的).

下面的定理给出了常返性和非常返性的*准则*.

定理 1 a) *状态 $i \in E$ 的常返性等价于如下两条性质中的任何一条:*

$$\mathbf{P}_i\{X_n = i \text{ 对于无限多个 } n\} = 1 \quad \text{或} \quad \sum_i p_{ii}^{(n)} = \infty.$$

b) *状态 $i \in E$ 的非常返性等价于如下两条性质中的任何一条:*

$$\mathbf{P}_i\{X_n = i \text{ 对于无限多个 } n\} = 0 \quad \text{或} \quad \sum_i p_{ii}^{(n)} < \infty.$$

于是, 由定理 1, 可见

$$f_{ii} = 1 \Leftrightarrow \mathbf{P}_i\{X_n = i \text{ 对于无限多个 } n\} = 1 \Leftrightarrow \sum_n p_{ii}^{(n)} = \infty, \tag{7}$$

$$f_{ii} < 1 \Leftrightarrow \mathbf{P}_i\{X_n = i \text{ 对于无限多个 } n\} = 0 \Leftrightarrow \sum_n p_{ii}^{(n)} < \infty. \tag{8}$$

注 注意, 根据第二章 §1 表 1, 事件 $\{X_n = i$ 对于无限多个 $n\}$, 是对于无限多个 n 使 $X_n(\omega) = i$ 的 ω 集合的. 如果这时 $A_n = \{\omega : X_n(\omega) = i\}$, 则 (见第二章 §1 表 1).

$$\{X_n = i \text{ 对于无限多个 } n\} = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

证明　立即可以指出如下蕴涵关系:

$$\sum_n p_{ii}^{(n)} < \infty \Rightarrow \mathbf{P}_i\{X_n = i \text{ 对于无限多个 } n\} = 0, \tag{9}$$

因为, 由于 $p_{ii}^{(n)} = \mathbf{P}_i\{X_n = i\}$, 可见上面的蕴含关系是博雷尔 – 坎泰利引理 1 的推论 (第二章 §10 命题 a)).

现在证明,

$$f_{ii} = 1 \Leftrightarrow \sum_n p_{ii}^{(n)} = \infty. \tag{10}$$

由齐次性和马尔可夫性, 可见对于任何数组 $(i_1, \cdots, i_k)$ 和 $(j_1, \cdots, j_n)$, 有

$$\begin{aligned}&\mathbf{P}_i\{(X_1, \cdots, X_k) = (i_1, \cdots, i_k), (X_{k+1}, \cdots, X_{k+n}) = (j_1, \cdots, j_n)\}\\ &= \mathbf{P}_i\{(X_1, \cdots, X_k) = (i_1, \cdots, i_k)\}\mathbf{P}_{i_k}\{(X_1, \cdots, X_n) = (j_1, \cdots, j_n)\}.\end{aligned}$$

由此直接得 (对照第一章 §12 公式 (38) 和第一章 §2 公式 (25) 的推导):

$$\begin{aligned}p_{ij}^{(n)} = \mathbf{P}_i\{X_n = j\} &= \sum_{k=0}^{n-1} \mathbf{P}_i\{X_1 \neq j, \cdots, X_{n-k-1} \neq j, X_{n-k} = j, X_n = j\}\\ &= \sum_{k=0}^{n-1} \mathbf{P}_i\{X_1 \neq j, \cdots, X_{n-k-1} \neq j, X_{n-k} = j\}\mathbf{P}_j\{X_k = j\}\\ &= \sum_{k=0}^{n-1} f_{ij}^{(n-k)} p_{jj}^{(k)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)}.\end{aligned}$$

这样, 得如下 $n(n \geqslant 1)$ 步转移概率的公式:

$$p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)}. \tag{11}$$

设 $j = i$, 有

$$\begin{aligned}\sum_{n=1}^{\infty} p_{ii}^{(n)} &= \sum_{n=1}^{\infty}\sum_{k=1}^{n} f_{ii}^{(k)} p_{ii}^{(n-k)} = \sum_{n=1}^{\infty} f_{ii}^{(k)} \sum_{n=k}^{\infty} p_{ii}^{(n-k)}\\ &= f_{ii}\sum_{n=0}^{\infty} p_{ii}^{(n)} = f_{ii}\left(1 + \sum_{n=1}^{\infty} p_{ii}^{(n)}\right),\end{aligned} \tag{12}$$

其中 $p_{ii}^{(0)} = 1$. 由此可见

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty \Rightarrow f_{ii} = \frac{\sum_{n=1}^{\infty} p_{ii}^{(n)}}{1 + \sum_{n=1}^{\infty} p_{ii}^{(n)}}. \tag{13}$$

现在设 $\sum\limits_{n=1}^{\infty} p_{ii}^{(n)} = \infty$, 则

$$\sum_{n=1}^{N} p_{ii}^{(n)} = \sum_{n=1}^{N}\sum_{k=1}^{n} f_{ii}^{(k)} p_{ii}^{(n-k)} = \sum_{k=1}^{N} f_{ii}^{(k)} \sum_{n=k}^{N} p_{ii}^{(n-k)} \leqslant \sum_{k=1}^{N} f_{ii}^{(k)} \sum_{l=0}^{N} p_{ii}^{(l)},$$

因而

$$f_{ii} = \sum_{k=1}^{\infty} f_{ii}^{(k)} \geqslant \sum_{k=1}^{N} f_{ii}^{(k)} \geqslant \frac{\sum\limits_{n=1}^{N} p_{ii}^{(n)}}{\sum\limits_{l=0}^{N} p_{ii}^{(l)}} \to 1, \quad N \to \infty.$$

从而

$$\sum_{n=1}^{N} p_{ii}^{(n)} = \infty \Rightarrow f_{ii} = 1. \tag{14}$$

由蕴含关系 (13) 和 (14), 立即得到互蕴涵关系:

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty \Leftrightarrow f_{ii} < 1, \tag{15}$$

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty \Leftrightarrow f_{ii} = 1. \tag{16}$$

为完成定理的证明, 只需验证下面的互蕴涵关系:

$$f_{ii} < 1 \Leftrightarrow \mathbf{P}_i\{X_n = i \text{ 对于无限多个 } n\} = 0, \tag{17}$$

$$f_{ii} = 1 \Leftrightarrow \mathbf{P}_i\{X_n = i \text{ 对于无限多个 } n\} = 1. \tag{18}$$

直观上, 这些性质是很清楚的. 例如, 如果 $f_{ii} = 1$, 则说明 $\mathbf{P}_i\{\sigma_i < \infty\} = 1$, 即 "质点" 迟早要返回它开始游动的状态 i. 但这是根据强马尔可夫性, 从这一 (随机) 时刻起, "质点的寿命" 仿佛重新开始. 假如继续观察, 则我们将看到, 事件 $\{X_n = i\}$ 将对于无限多个下标 n 出现, 即 $\mathbf{P}_i\{X_n = i, \text{ 对于无限多个 } n\} = 1$.

我们现在对性质 (17) 和 (18) 进行严格证明.

对于给定的状态 $i \in E$, 考虑回返状态i 的次数不少于 m 的概率 α_m. 我们证明此概率等于 $\alpha_m = (f_{ii})^m$.

事实上, 如果 $m = 1$, 则 f_{ii} 由定义可知 $\alpha_1 = f_{ii}$. 现在假设 $\alpha_{m-1} = (f_{ii})^{m-1}$, 证明 $\alpha_m = (f_{ii})^m$.

由强马尔可夫性 (见 §2 的 (8) 式), 并注意到事件 $\{\sigma_i = k\} \in \mathscr{F}_{\sigma_i}$, 可见

$$
\begin{aligned}
\alpha_m &= \mathbf{P}_i\{\text{返回状态 } i \text{ 的次数不少于 } m\} \\
&= \sum_{k=1}^{\infty} \mathbf{P}_i\{\sigma_i = k \text{ 且在 } k \text{ 时后返回状态 } i \text{ 的次数不少于 } m-1\} \\
&= \sum_{k=1}^{\infty} \mathbf{P}_i\{\sigma_i = k\}\mathbf{P}_i\{X_{\sigma_i+1}, X_{\sigma_i+2}, \cdots \text{ 中至少有 } m-1 \text{ 个值等于 } i | \sigma_i = k\} \\
&= \sum_{k=1}^{\infty} \mathbf{P}_i\{\sigma_i = k\}\mathbf{P}_i\{X_1, X_2, \cdots \text{ 中至少有 } m-1 \text{ 个值等于 } i\} \\
&= \sum_{k=1}^{\infty} f_{ii}^{(k)}(f_{ii})^{m-1} = f_{ii}(f_{ii})^{m-1} = (f_{ii})^m.
\end{aligned}
$$

由此得

$$
\mathbf{P}_i\{X_n = i \text{ 对于无限多个 } n\} = \lim_{m\to\infty}(f_{ii})^m = \begin{cases} 1, & \text{若 } f_{ii} = 1, \\ 0, & \text{若 } f_{ii} = 0. \end{cases} \tag{19}
$$

该公式表明, 如果 $A = \{A_n \text{ 对于无限多个 } n\}$ $(= \overline{\lim} A_n)$, 其中 $A_n = \{X_n = i\}$, 则对于概率 $\mathbf{P}_i(A)$, "0 – 1 律" 成立, 即 $\mathbf{P}_i(A)$ 只有 0 和 1 两个可能值. 注意, 这一性质不能直接由博雷尔 – 坎泰利引理 (第二章 §10) 直接得到, 因为事件 $A_n (n \geqslant 1)$ 一般不独立.

由 (19) 式, 以及性质: $\mathbf{P}_i(A)$ 只有 0 和 1 两个可能值, 可得所要证明的 (17) 式和 (18) 式中的蕴涵关系. □

2. 非常返状态 由定理 1 可以得到非常返状态如下简单而重要的性质.

定理 2 如果状态 j 非常返, 则对于任意状态 $i \in E$,

$$
\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty, \tag{20}
$$

因此, 任意状态 $i \in E$,

$$
p_{ij}^{(n)} \to 0, \quad n \to \infty. \tag{21}
$$

证明 由 (11) 式 (其中 $p_{jj}^{(0)} = 1$), 有

$$
\begin{aligned}
\sum_{n=1}^{\infty} p_{ij}^{(n)} &= \sum_{n=1}^{\infty}\sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)} = \sum_{k=1}^{\infty} f_{ij}^{(k)} \sum_{n=0}^{\infty} p_{jj}^{(n)} \\
&= f_{ij}\sum_{n=0}^{\infty} p_{jj}^{(n)} \leqslant \sum_{n=0}^{\infty} p_{jj}^{(n)} < \infty,
\end{aligned}
$$

其中我们注意到

$$
f_{ij} = \sum_{k=1}^{\infty} f_{ij}^{(k)} \leqslant 1
$$

(f_{ij} 是质点从状态 i 出发, 迟早到达状态 j 的概率).

由 (20) 式显然得 (21) 式. □

3. 常返状态 现在考虑常返状态. 以首返某状态 $i \in E$ 的平均首返时间

$$\mu_i = \sum_{n=1}^{\infty} f_{ii}^{(n)} \leqslant 1 \quad (= \mathbf{E}_i \sigma_i) \tag{22}$$

有限还是无限为转移, 每一个常返状态 $i \in E$ 分为两种类型: 正常返或零常返状态. 我们知道, 根据 (1) 式 $f_{ii}^{(n)}$ 表示恰好经过 n 步首返的概率.

定义 3 如果

$$\mu_i^{-1} = \left(\sum_{n=1}^{\infty} n f_{ii}^{(n)}\right)^{-1} > 0, \tag{23}$$

则称状态 $i \in E$ 为正常返的, 而如果

$$\mu_i^{-1} = \left(\sum_{n=1}^{\infty} n f_{ii}^{(n)}\right)^{-1} = 0, \tag{24}$$

则称状态 $i \in E$ 为零常返的.

这样, 根据定义首返零常返状态, 平均经过无限时间出现. 平均首返正常返状态的时间是有限的.

4. 马尔可夫链的状态按转移概率矩阵的渐近性质分类 下面的框图直观地表示, 马尔可夫链的状态, 根据常返和非常返, 正常返和零常返分类.

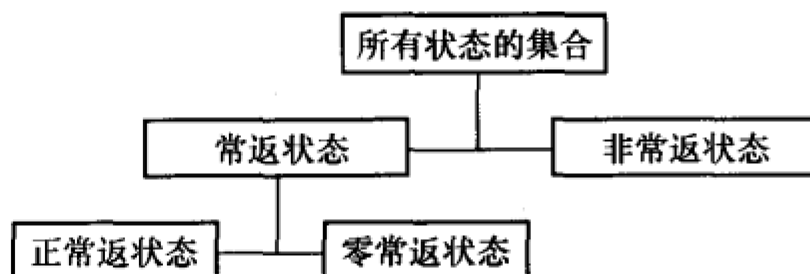


图 40 马尔可夫链的状态按概率 $p_{ij}^{(n)}$ 的渐近性质分类

5. 常返和非周期状态下转移概率矩阵的渐近性质

定理 3 设马尔可夫链的状态 $i \in E$ 是常返的和非周期的 $(d(j) = 1)$. 那么, 对于任意 $i \in E$, 有

$$p_{ij}^{(n)} \to \frac{f_{ij}}{\mu_j}, \quad n \to \infty. \tag{25}$$

此外, 如果状态 i 和 j 是互通的 $(i \leftrightarrow j)$, 即状态 i 和 j 属于同一不可约类, 则

$$p_{ij}^{(n)} \to \frac{1}{\mu_j}, \quad n \to \infty. \tag{26}$$

在下面将要进行的证明中, 本质地依赖于引理 1 的论点, 而该论点是 "离散更新理论" 的关键结果. 关于定理 3 的另外一个, 基于耦合 (coupling, 第三章 §8) 的思想, 例如参见 [104], [105].

引理 1 ("离散更新理论" 的基本引理)　设 $\varphi = (\varphi_1, \varphi_2, \cdots)$ 是非负数的非周期 $(d(j) = 1)$ 序列, 根据该序列建立一对应于如下递推规则的序列 $u = (u_1, u_2, \cdots)$: $u_0 = 1$, 对于任意 $n \geqslant 1$,

$$u_n = \varphi_1 u_{n-1} + \varphi_2 u_{n-2} + \cdots + \varphi_n u_0. \tag{27}$$

那么, 当 $n \to \infty$ 时

$$u_n \to \mu^{-1},$$

其中 $\mu = \sum\limits_{n=1}^{\infty} n\varphi_n$.

关于该引理的证明, 例如参见 [69] (第 1 卷, 第 VIII 章, 10).

定理 3 的证明　首先设 $i = j$. 证明

$$p_{ij}^{(n)} \to \frac{1}{\mu_j}, \quad n \to \infty. \tag{28}$$

为此将 (11) 式 (对于 $i = j$ 的情形) 写成下面的形式:

$$p_{jj}^{(n)} = f_{jj}^{(1)} p_{jj}^{(n-1)} + f_{jj}^{(2)} p_{jj}^{(n-2)} + \cdots + f_{jj}^{(n)} p_{jj}^{(0)}, \tag{29}$$

其中设 $p_{jj}^{(0)} = 1$, 并且显然 $f_{jj}^{(1)} = p_{jj}^{(1)}$. 如果设

$$u_k = p_{jj}^{(k)}, \quad \varphi_k = f_{jj}^{(k)}, \tag{30}$$

则 (29) 式有如下形式

$$u_n = \varphi_1 u_{n-1} + \varphi_2 u_{n-2} + \cdots + \varphi_n u_0,$$

因此恰好是引理 1 的递推公式.

假如我们能证明 (定理 3 中假设的条件) 序列 $(p_{jj}^{(1)}, p_{jj}^{(2)}, \cdots)$ 的周期等于 1, 则序列 $(f_{jj}^{(1)}, f_{jj}^{(2)}, \cdots)$ 的周期 $d_f(j)$ 等于 1, 那么所要证明的 (28) 式的结果可以直接由引理 1 得到.

而这又是下面 (引理 2 的) 一般命题的推论.

引理 2　对于任意 $j \in E$,

$$\mathrm{GCD}\{n \geqslant 1 : p_{jj}^{(n)} > 0\} = \mathrm{GCD}\{n \geqslant 1 : f_{jj}^{(n)} > 0\}, \tag{31}$$

即周期 $d_f(j) = d(j)$.

证明 设

$$M = \{n : p_{jj}^{(n)} > 0\} \quad 和 \quad M_f = \{n : f_{jj}^{(n)} > 0\}.$$

由于 $M_f \subseteq M$, 可见

$$\mathrm{GCD}(M) \leqslant \mathrm{GCD}(M_f),$$

即 $d(j) \leqslant d_f(j)$.

相反的不等式由如下概率的意义 $p_{jj}^{(n)}$ 和 $f_{jj}^{(n)}, n \geqslant 1$ 可以得到.

如果 "质点" 从状态 j 出发经过 n 步仍然处于该状态 $(p_{jj}^{(n)} > 0)$, 则说明其游动是这样进行的: 首先经 k_1 步 $(f_{jj}^{(k_1)} > 0)$ 由状态 j 首次返回 j, 然后经 k_2 步 $(f_{jj}^{(k_2)} > 0), \cdots$, 经 k_l 步 $(f_{jj}^{(k_l)} > 0)$ 由状态 j 首次返回 j.

因而 $n = k_1 + k_2 + \cdots + k_l$. 数 $d_f(j)$ 整除 $k_1, k_2, \cdots, k_l$, 故也整除 n. 由于 $d(j)$ 是被 n 整除且使 $p_{jj}^{(n)} > 0$ 的数中最大者, 可见 $d(j) \geqslant d_f(j)$.

于是 $d(j) = d_f(j)$. 顺便指出, 在利用公式 $d(j) = \mathrm{GCD}\{n \geqslant 1 : p_{jj}^{(n)} > 0\}$ 定义状态 j 的周期 $d(j)$ 时, 也可以利用公式 $d(j) = \mathrm{GCD}\{n \geqslant 1 : f_{jj}^{(n)} > 0\}$. 引理 2 得证. □

完成定理 3 的证明 1) (25) 式的证明 ($i \neq j$ 的情形).

将 (11) 式表示为如下形式:

$$p_{ij}^{(n)} = \sum_{k=1}^{\infty} f_{ij}^{(k)} p_{jj}^{(n-k)}, \tag{32}$$

其中 $p_{jj}^{(l)} = 0, l < 0$.

这里, 由于当 $n \to \infty$ 时

$$p_{ij}^{(n)} \to \frac{1}{\mu_j}, \quad 且 \quad \sum_{k=1}^{\infty} f_{ij}^{(k)} \leqslant 1,$$

可见根据控制收敛定理 (第二章 §6 定理 3)

$$\lim_n \sum_{k=1}^{\infty} f_{ij}^{(k)} p_{jj}^{(n-k)} = \sum_{k=1}^{\infty} f_{ij}^{(k)} \lim_n p_{jj}^{(n-k)} = \frac{1}{\mu_j} \sum_{k=1}^{\infty} f_{ij}^{(k)} = \frac{f_{ij}}{\mu_j}. \tag{33}$$

由 (32) 和 (33) 式得

$$\lim_n p_{ij}^{(n)} = \frac{f_{ij}}{\mu_j}, \tag{34}$$

即命题 (25) 得证.

2) (26) 式的证明 ($i \leftrightarrow j$ 的情形). 假如在条件 $i \leftrightarrow j$ 下 (即状态 i 和 j 属于同一互通状态的不可约类) 概率 $f_{ij} = 1$, 则由 (34) 式即可得性质 (26) 式.

假设状态 j 是常返的, 则由定理 1 的命题 a), 有 $\mathbf{P}_j\{X_n = j$ 对于无限多个 $n\} = 1$. 因此, 对于任意 m,

$$\begin{aligned} p_{ji}^{(m)} &= \mathbf{P}_j(\{X_m = i\} \bigcap \{X_n = j \text{ 对于无限多个 } n\}) \\ &\leqslant \sum_{n>m} \mathbf{P}_j\{X_m = i, X_{m+1} \neq j, \cdots, X_{n-1} \neq j, X_n = j\} \\ &= \sum_{n>m} p_{ji}^{(m)} f_{ij}^{(n-m)} = p_{ji}^{(m)} f_{ij}, \end{aligned} \tag{35}$$

其中中间的不等式是广义马尔可夫性的推论 (见 §2 的 (2) 式). 由于 E 是互通状态类, 可见存在 m, 使 $p_{ji}^{(m)} > 0$. 因此由 (35) 式推得 $f_{ij} = 1$. □

6. 状态周期任意的情形 对于感兴趣的状态 j 的周期 d 任意 $(d = d(j) > 0)$ 的情形, 可以自然地表述上面的定理 3 的类似. 看下面的定理.

定理 4 *设马尔可夫链的状态 $j \in E$ 是常返, 且其周期的 $d = d(j) \geqslant 1$, 而 $i \in E$ 也是马尔可夫链的状态 j (有可能与状态 j 相同).*

a) 设 i 和 j 属于同一不可分解状态类 $C \subseteq E$, 且 $C_0, C_1, \cdots, C_{d-1}$ 是 (循环的) 子类, 按如下顺序编号: $j \in C_0, i \in C_a$ 其中 $a \in \{0, 1, \cdots, d-1\}$, 而在此子类中的运动是按如下顺序: $C_0 \to C_1 \to \cdots \to C_a \to \cdots \to C_{d-1} \to C_0$. 那么, 当 $n \to \infty$ 时

$$p_{ij}^{(nd+a)} \to \frac{d}{\mu_j}, \tag{36}$$

b) 在一般情形下, 当 i 和 j 可能属于不同的不可约状态类时, 当 $n \to \infty$ 时, 对于任意 $a = 0, 1, \cdots, d-1$,

$$p_{ij}^{(nd+a)} \to \frac{d}{\mu_j} \sum_{k=0}^{n} f_{jj}^{(kd+a)}. \tag{37}$$

证明 首先设 $a = 0$, 即 i 和 j 可能属于同一不可约状态类 C, 并且同一循环子类 C_0.

考虑转移概率 $p_{ij}^{(d)}, i, j \in C$, 并根据 $p_{ij}^{(d)}$ (按照 §1 的构造) 建立新的马尔可夫链. 对于新链, 状态 j 是常返的和非周期的. 新链的状态 i 和 j 仍然是互通的 $(i \leftrightarrow j)$. 这就证明了定理 3 的性质 (26):

$$p_{ij}^{(nd)} \to \frac{1}{\sum_{k=1}^{\infty} k f_{jj}^{(kd)}} = \frac{d}{\sum_{k=1}^{\infty} (kd) f_{jj}^{(kd)}} = \frac{d}{\mu_j},$$

其中最后一等式成立, 因为, 对于一切不能被 d 整除的 l, 有 $f_{jj}^{(l)} = 0$, 并且根据定义.

$$\mu_j = \sum_{l=1}^{\infty} l f_{jj}^{(l)}.$$

现在假设, 对于 $a=0,1,\cdots,r(\leqslant d-2)$, (36) 式已得证.

根据控制收敛定理 (第二章 §6 定理 3),

$$p_{ij}^{(nd+r+1)}=\sum_{k=1}^{\infty}p_{ik}p_{kj}^{(nd+r)}\to\sum_{k=1}^{\infty}p_{ik}\frac{d}{\mu_j}=\frac{d}{\mu_j}.$$

于是, 对于 $a=r+1(\leqslant d-1)$, 欲证明的 (36) 式得证. 从而, 根据归纳法, 对于一切 $a=0,1,\cdots,d-1$, (36) 式得证.

b) 对于任何 $i,j\in E$, 下面的公式成立 (见 (11) 式):

$$p_{ij}^{(nd+a)}=\sum_{k=1}^{nd+a}f_{ij}^{(k)}p_{jj}^{(nd+a-k)},\quad a=0,1,\cdots,d-1.$$

根据假设状态 j 的周期等于 d. 因此, 除当 $k-a$ 具有形式 rd 时之外, 有 $p_{jj}^{(nd+a-k)}=0$. 于是, 有

$$p_{ij}^{(nd+a)}=\sum_{r=0}^{n}f_{ij}^{(rd+a)}p_{jj}^{((n-r)d)}.$$

由此以及上面证明的 (36) 式, 仍然运用控制收敛定理, 最后得到需要证明的公式 (37). □

7. 非周期马尔可夫链的完全分类 像在 §4 的末尾指出的那样, 在根据转移概率的渐近性质, 研究马尔可夫链的状态的分类时, 只需局限于考虑*非周期不可约链*.

在定理 1~3 中陈述的结果, 实际上包含了这样链的完全分类的全部必要内容.

首先引进一个辅助命题, 它是关于不可约链的所有状态属于同一种 ("常返" 或 "非常返") 类型. (对照 §4 定理 2 中的 "单一类型性".)

引理 3 *设 E 是 (互通状态的) 的不可约类. 那么, 其全部状态或者都是常返的, 或者都是非常返的.*

证明 假设链至少有一个非常返状态, 例如状态 i, 则根据定理 1, $\sum_n p_{ii}^{(n)}<\infty$.

现在假设是 j 任意另一个状态. 由于 E 互通状态 $(i\leftrightarrow j)$ 的不可约类, 存在这样的状态 k 和 l, 使 $p_{ij}^{(k)}>0$ 和 $p_{ij}^{(l)}>0$. 那么, 由明显的不等式

$$p_{ii}^{(n+k+l)}\geqslant p_{ij}^{(k)}p_{jj}^{(n)}p_{ji}^{(l)},$$

可见

$$\sum_n p_{ii}^{(n+k+l)}\geqslant p_{ij}^{(k)}p_{ji}^{(l)}\sum_n p_{jj}^{(n)}.$$

根据假设 $\sum_n p_{ii}^{(n)}<\infty$, 而 k,l 满足 $p_{ij}^{(k)}p_{ji}^{(l)}>0$, 从而 $\sum_n p_{jj}^{(n)}<\infty$.

鉴于定理 1 的 b), 由此可见, 状态 j 也是非常返状态. 换句话说, 假如不可约类只要有一个非常返状态, 则其余状态也都是非常返状态.

现在设 i 是常返状态, 我们证明其余状态也都是常返状态.

假设 (除 i 之外) 至少有一个非常返状态. 那么根据上面已证明的结果, 其余状态也都是非常返状态. 因此与 "i 是常返状态" 的假设矛盾.

因此哪怕有一个常返状态, 自动导致 (不可约链的) 全部状态都是常返状态. 从而引理得证. □

该引理的结论, 完全证明了对于不可约链,"常返链" 和 "非常返链" 这两个 (被普遍接受的) 术语是合适的. (注意, 这里说的 "链", 而不仅是个别状态.)

定理 5 假设马尔可夫链由非周期状态的一个不可约类 E 构成. 对于这样的链仅属于如下三种可能的类型之一:

(i) 非常返链. 在这种情形下, 对于一切 $i, j \in E$,

$$\lim_n p_{ij}^{(n)} = 0,$$

并且在

$$\sum_n p_{ij}^{(n)} < \infty$$

的意义上充分 "快" 地收敛于 0.

(ii) 常返与零常返链. 在这种情形下, 对于一切 $i, j \in E$,

$$\lim_n p_{ij}^{(n)} = 0,$$

并且收敛在

$$\sum_n p_{ij}^{(n)} = \infty$$

的意义上充分地 "慢", 此外从 j 到 j 的平均首返时间 μ_j 等于 ∞.

(iii) 常返与正常返链. 在这种情形下, 对于一切 $i, j \in E$,

$$\lim_n p_{ij}^{(n)} = \frac{1}{\mu_j} > 0,$$

其中 μ_j 是从 j 到 j 的平均首返时间, 并且有限.

证明 命题 (i) 在定理 1 的 b) 和定理 2 已经证明. 命题 (ii) 和 (iii) 可以直接由定理 1 的 a) 和定理 3 得到. □

8. 有限马尔可夫链 现在考虑有限马尔可夫链的情形, 即状态集合 E 由有限个元素组成的情形.

结果, 在这种情形下, 定理 5 中的三种可能性 (i), (ii), (iii) 只有 (iii) 仍然成立.

定理 6 设有限马尔可夫链是不可约的和非周期的. 那么, 这样的链是常返的和正的. 这时

$$\lim_n p_{ij}^{(n)} = \frac{1}{\mu_j} > 0.$$

证明 用反证法. 假设此链是非常返的. 那么, 如果链的状态数 r 是有限的 $(E=1,2,\cdots,r)$, 则

$$\lim_n \sum_{j=1}^{r} p_{ij}^{(n)} = \sum_{j=1}^{r} \lim_n p_{ij}^{(n)}. \tag{38}$$

该式的右侧显然等于 1. 但是我们假设它是非常返的, 那么, 这时 (根据定理 5 的 (i)) 右侧应该等于 0. 由此得到的矛盾将 (i) 排除.

现在假设链是常返的.

由于定理 5 的命题只剩下两种可能性 (ii) 和 (iii). 需要排除 (ii). 事实上, 由于对于所有 $i,j\in E, \lim_n p_{ij}^{(n)}=0$, 则像非常返状态的情形一样, 出现矛盾.

于是, 只有第三种可能性 (iii). □

9. 练习题

1. 考虑不可约链, 其状态集为 $0,1,2,\cdots$ 那么, 它是非常返的, 当且仅当方程组

$$u_j = \sum_i u_i p_{ij}, \quad j=0,1,\cdots$$

有有界解, 而且 $u_i \not\equiv c$ (常数), $i=0,1,\cdots$

2. 只要序列 $u=(u_0,u_1,\cdots), u_i\to\infty, i\to\infty$, 使得对于一切 $j\neq 0, u_i \geqslant \sum\limits_i u_i p_{ij}$, 则状态集为 $0,1,2,\cdots$ 的不可约链是常返的.

3. 状态集为 $0,1,2,\cdots$ 的不可约链是常返的和正的, 当且仅当方程组

$$u_j = \sum_i u_i p_{ij}, \quad j=0,1,\cdots$$

有不恒等于 0 的解, 而且 $\sum\limits_i |u_i| < \infty$.

4. 假设马尔可夫链的状态为 $0,1,\cdots$, 而转移概率为

$$p_{00}=r_0, \quad p_{01}=p_0>0,$$
$$p_{ij}=\begin{cases} p_i>0, & j=i+1,\\ r_i\geqslant 0, & j=i,\\ q_i>0, & j=i-1,\\ 0, & \text{其他}.\end{cases}$$

设

$$\rho_0=1, \quad \rho_m=\frac{q_1\cdots q_m}{p_1\cdots p_m}.$$

证明如下的等价关系:

$$\text{常返链} \Leftrightarrow \sum \rho_m = \infty,$$
$$\text{非常返链} \Leftrightarrow \sum \rho_m < \infty,$$
$$\text{正链} \Leftrightarrow \sum \rho_m = \infty, \quad \sum \frac{1}{p_m \rho_m} < \infty,$$
$$\text{零链} \Leftrightarrow \sum \rho_m = \infty, \quad \sum \frac{1}{p_m \rho_m} = \infty.$$

5. 证明:

$$f_{ik} \geqslant f_{ij} f_{jk},$$
$$\sup_n p_{ij}^{(n)} \leqslant f_{ij} \leqslant \sum_{n=1}^{\infty} p_{ij}^{(n)}.$$

6. 证明, 对于任意具有可数状态集的马尔可夫链, 对于转移概率 $p_{ij}^{(n)}$, 在切萨罗 (E. Cesàro) 意义上永远存在极限:

$$\lim_n \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)} = \frac{f_{ij}}{\mu_j}.$$

7. 考虑马尔可夫链 $\xi_0, \xi_1, \cdots$, 且 $\xi_{k+1} = (\xi_k)^+ + \eta_{k+1}, k \geqslant 0$, 其中 $\eta_1, \eta_2, \cdots$ 是独立同分布随机变量序列, 其概率分布为: $\mathbf{P}\{\eta_k = j\} = p_j, j = 0, 1, \cdots$

(a) 试写出转移概率矩阵;

(b) 证明, 如果 $p_0 > 0, p_0 + p_1 < 1$, 则链常返当且仅当

$$\sum_k k p_k \leqslant 1.$$

§6. 可数马尔可夫链的极限分布、遍历分布和平稳分布

1. 极限值 $\mathbb{\Pi}$ 与平稳分布 $\mathbb{Q}$ 的联系 我们从一个一般结果开始, 它首先与非常清晰的极限值 $\mathbb{\Pi} = (\pi_1, \pi_2, \cdots)$ 相联系, 其中 $\pi_j = \lim_n p_{ij}^{(n)} (j = 1, 2, \cdots)$, 以及与平稳分布 $\mathbb{Q} = (q_1, q_2, \cdots)$ 相联系.

定理 1 *考虑具有可数状态集 $E = \{1, 2, \cdots\}$ 的马尔可夫链, 其转移概率 $p_{ij}^{(n)} (i, j \in E)$ 有不依赖于初始状态 $i \in E$ 的极限:*

$$\pi_j = \lim_n p_{ij}^{(n)}, \quad j \in E.$$

那么,

(a) $\sum_{j=1}^{\infty} \pi_j \leqslant 1, \sum_{j=1}^{\infty} \pi_i p_{ij} = \pi_j, j \in E;$

(b) 下面的式子二者必居其一:

$$\sum_{j=1}^{\infty}\pi_j = 0(\text{ 即所有 } \pi_j = 0, j \in E) \quad \text{或} \quad \sum_{j=1}^{\infty}\pi_j = 1;$$

(c) 如果 $\sum\limits_{k=1}^{\infty}\pi_j = 0$, 则马尔可夫链无平稳分布; 如果 $\sum\limits_{j=1}^{\infty}\pi_j = 1$, 则极限值的向量 $\Pi = (\pi_1, \pi_2, \cdots)$, 就是马尔可夫链的平稳分布, 而且该马尔可夫链再无其他平稳分布.

证明 有

$$\sum_{j=1}^{\infty}\pi_j = \sum_{j=1}^{\infty}\lim_n p_{ij}^{(n)} \leqslant \varliminf_n \sum_{j=1}^{\infty} p_{ij}^{(n)} = 1, \tag{1}$$

而对于任意 $j \in E, k \in E$, 有

$$\sum_{i=1}^{\infty}\pi_i p_{ij} = \sum_{i=1}^{\infty}\lim_n p_{ki}^{(n)} p_{ij} \leqslant \varliminf_n \sum_{i=1}^{\infty} p_{ki}^{(n)} p_{ij} = \varliminf_n p_{kj}^{(n+1)} = \pi_j. \tag{2}$$

注 这里出现的不等式和下极限, 当然是"法图引理"的推论. 不过应该注意, 在这里法图引理, 并不是像第二章 §6 那样, 用于由概率测度定义的勒贝格积分, 而是用于按 $\sigma-$ 有限 (非负) 测度进行积分的情形.

这样, 极限值的向量 $\Pi = (\pi_1, \pi_2, \cdots)$, 具有如下性质:

$$\sum_{j=1}^{\infty}\pi_j \leqslant 1, \quad \sum_{i=1}^{\infty}\pi_i p_{ij} \leqslant \pi_j, \quad j \in E. \tag{3}$$

现在证明, 最后一个不等式实际上是等式.

设对于任意 $j_0 \in E$, 有

$$\sum_{i=1}^{\infty}\pi_i p_{ij_0} \leqslant \pi_{j_0}. \tag{4}$$

那么,

$$\sum_{j=1}^{\infty}\pi_j > \sum_{j=1}^{\infty}\left(\sum_{i=1}^{\infty}\pi_i p_{ij}\right) = \sum_{i=1}^{\infty}\pi_i \sum_{j=1}^{\infty} p_{ij} = \sum_{i=1}^{\infty}\pi_i.$$

所得矛盾说明, (3) 式中最后一个不等式实际上是等式. 由于对于 $j \in E$, 有

$$\sum_{j=1}^{\infty}\pi_j \leqslant 1, \quad \sum_{i=1}^{\infty}\pi_i p_{ij} = \pi_j,$$

可见性质 (a) 成立.

为证明性质 (b), 注意到 (3) 式 $\sum\limits_{i=1}^{\infty}\pi_i p_{ij} = \pi_j$, 经积分, 对于任意 $n \geqslant 1$ 和任意 $j \in E$, 得

$$\sum_{i=1}^{\infty}\pi_i p_{ij}^{(n)} = \pi_j.$$

根据勒贝格控制收敛定理 (第二章 §6 定理 3), 由此可见

$$\pi_j = \lim_n \sum_{i=1}^{\infty} \pi_i p_{ij}^{(n)} = \sum_{j=1}^{\infty} \pi_i \lim_n p_{ij}^{(n)} = \left(\sum_{i=1}^{\infty} \pi_i\right) \pi_j,$$

即

$$\pi_j \left(1 - \sum_{i=1}^{\infty} \pi_i\right) = 0, \quad j \in E,$$

因此

$$\left(\sum_{j=1}^{\infty} \pi_j\right) \left(1 - \sum_{i=1}^{\infty} \pi_i\right) = 0.$$

这样, $a(1-a)=0$, 而 $a = \sum_{i=1}^{\infty} \pi_i$. 所以 $a=1$ 或 $a=0$, 这就证明了命题 (b).

最后证明 (c). 为此, 假设 $\mathbb{Q} = (q_1, q_2, \cdots)$ 是某一平稳分布, 则

$$\sum_{i=1}^{\infty} q_i p_{ij}^{(n)} = q_j, \quad 且 \quad \left(\sum_{i=1}^{\infty} q_i\right) \pi_j = q_j, \quad j \in E,$$

其中后一个等式成立, 是根据控制收敛定理.

所以, 如果 $\mathbb{Q}$ 是平稳分布, 则 $\sum_{i=1}^{\infty} q_i = 1$, 从而对于此平稳分布一定 (应该) 满足条件: 对于一切 $j \in E$, 有 $q_j = \pi_j$. 这样一来, 假如 $\sum_{j=1}^{\infty} \pi_j = 0$, 则性质 $\sum_{i=1}^{\infty} q_i = 1$ 就不可能成立, 说明在这种情形下没有平稳分布.

根据 (b) 还有一种可能性 $\sum_{j=1}^{\infty} \pi_j = 1$. 在这种情形下, 根据 (a), $\mathbb{\Pi} = (\pi_1, \pi_2, \cdots)$ 本身是平稳分布, 而由以上的叙述知, 如果 $\mathbb{Q}$ 是另外某个平稳分布, 则它应与 $\mathbb{\Pi}$ 重合. 于是, 在 $\sum_{j=1}^{\infty} \pi_j = 1$ 的情形下, 平稳分布的唯一性也得到证明. □

2. 平稳分布和遍历分布的基本定理　定理 1 给了平稳分布存在 (同时也唯一) 的*充分*条件. 该条件在于对于一切 $j \in E$, *存在*不依赖于 $i \in E$ 的极限值 $\pi_j = \lim_n p_{ij}^{(n)}$, 并且至少对于一个状态 $j \in E$, 使得 $\pi_j > 0$.

同时, 在 §5 中相当详细地研究了极限 $\lim_n p_{ij}^{(n)}$ *存在*的更为一般的问题, 与如下一些链的 "内在" 性质相联系, 诸如不可约性, 周期性, 常返性和非常返性, 正常返性和零常返性. 所以, 自然正是用这些由转移概率矩阵 $p_{ij}(i, j \in E)$ 决定的、"内在" 性质的术语, 来表述平稳分布存在的条件. 显然同样地, 如果用这些术语指出一切极限值 $\pi_j > 0 (j \in E)$ 的条件, 则根据定义 (见 §3 的性质 **C**) 向量 $\mathbb{\Pi} = (\pi_1, \pi_2, \cdots)$ 将形成*遍历*极限分布.

下面的两个定理将给出这些问题的答案.

定理 2 (平稳分布的基本定理) 考虑具有可数状态集 E 的马尔可夫链. 存在唯一平稳分布的必要和充分条件是:

(a) 恰好存在一个不可约子类,

(b) 一切状态都是正常返的.

定理 3 (遍历分布的基本定理) 考虑具有可数状态集的马尔可夫链. 存在遍历分布的必要和充分条件: 链是

(a) 不可约的,

(b) 正常返的,

(c) 非周期的.

3. 定理 2 的证明 必要性. 假设所考虑的马尔可夫链有唯一平稳分布, 记作 $\mathbb{Q}_s$. 那么, 我们证明, 在状态集 E 上存在并且唯一正常返子类.

以 $N(0 \leqslant N \leqslant \infty)$ 表示这样子类的潜在可能数.

设 $N = 0$, 而 j 是 E 中的某个状态. 由于不存在正常返类, 则状态 j 可能或者是非常返的, 或者是零常返的.

在第一种情形中, 由 §5 的定理 2 可见, 对于 $i \in E$, 极限 $\pi_j = \lim\limits_n p_{ij}^{(n)}$ 存在, 并且等于 0.

而在第二种情形中, 由 §5 的性质 (37) 以及由于 $\mu_j = \infty$ (因为状态 j 是零常返的) 可见, 这些极限 $\pi_j = \lim\limits_n p_{ij}^{(n)}$ 也存在, 并且等于 0.

这样, 当 $N = 0$ 时对于一切 $i, j \in E$, 极限 $\pi_j = \lim\limits_n p_{ij}^{(n)}$ 存在, 并且等于 0. 因此在这种情形下, 根据定理 1 的命题 (c), 在这种情形下没有平稳分布, 从而 $N = 0$ 的情形被 "存在平稳分布 $\widetilde{\mathbb{Q}}_s$" 的假设排除.

现在假设 $N = 1$. 以 C 表示唯一正常返类.

如果该类的周期 $d(C) = 1$, 则根据 §5 定理 3 的性质 (26), 可见对于一切 $i, j \in C$,

$$p_{ij}^{(n)} \to \mu_j^{-1}, \quad n \to \infty.$$

而若 $j \notin C$, 则这一状态非常返. 那么, 根据 §5 定理 2 的性质 (21), 可见对于一切 $i \in E$, 有

$$p_{ij}^{(n)} \to 0, \quad n \to \infty.$$

设

$$q_j = \begin{cases} \mu_j^{-1}(>0), & \text{若 } j \in C. \\ 0, & \text{若 } j \notin C. \end{cases} \tag{5}$$

那么, 由于集合 $C \neq \varnothing$, 则根据定理 1 数组 $\mathbb{Q} = (q_1, q_2, \cdots)$ 是唯一平稳分布, 从而 $\mathbb{Q} = \widetilde{\mathbb{Q}}_s$.

现在假设周期 $d(C) > 1$.

设 $C_0, C_1, \cdots, C_{d-1}$ 是 (正常返) 类的 C 的循环子类.

每一个子类 $C_k, k=0,1,\cdots,d-1$, 关于转移概率矩阵 $p_{ij}^{(d)}, i,j\in C$, 是常返的和非周期的. 那么, 如果 $i,j\in C_k$, 则根据 §5 公式 (36), 有

$$p_{ij}^{(nd)} \to \frac{d}{\mu_j} > 0.$$

所以在每一个集合 C_k 上, (由于定理 1 的性质 (b)) 数组 $\{d/\mu_j, j\in C_k\}$ (关于矩阵 $p_{ij}^{(d)}, i,j\in C$) 是唯一平稳分布.

特别, 由此可见

$$\sum_{j\in C_k}\frac{d}{\mu_j}=1, \quad 即 \quad \sum_{j\in C_k}\frac{1}{\mu_j}=\frac{1}{d}.$$

设

$$q_j=\begin{cases}\mu_j^{-1}, j\in C=C_0+\cdots+C_{d-1},\\ 0, \quad j\notin C.\end{cases} \tag{6}$$

因而, 就证明了, 对于所考虑的马尔可夫链, 数组 $\mathbb{Q}=(q_1,q_2,\cdots)$ 是唯一平稳分布.

事实上, 如果 $i\in C$, 则

$$p_{ii}^{(nd)}=\sum_{j\in C}P_{ij}^{(nd-1)}p_{ji}.$$

那么, 如同 (1) 式, 有

$$\frac{d}{\mu_i}=\lim_n p_{ii}^{(nd)} \geqslant \sum_{j\in C}\varliminf_n p_{ij}^{(nd-1)}p_{ji}=\sum_{j\in C}\frac{d}{\mu_j}p_{ji},$$

从而

$$\frac{1}{\mu_i}\geqslant\sum_{j\in C}\frac{1}{\mu_j}p_{ji}. \tag{7}$$

但是

$$\sum_{i\in C}\frac{1}{\mu_i}=\sum_{k=0}^{d-1}\left(\sum_{i\in C_k}\frac{1}{\mu_i}\right)=\sum_{k=0}^{d-1}\frac{1}{d}=1. \tag{8}$$

像定理 1 的证明 (见 (3) 式和 (4) 式) 一样, 由 (7) 式和 (8) 式, 可见 (7) 式中实际上是等式:

$$\frac{1}{\mu_i}=\sum_{j\in C}\frac{1}{\mu_j}p_{ji}. \tag{9}$$

由于 $q_i=\mu_i^{-1}>0$, 则 (9) 式表示数组 $\mathbb{Q}=(q_1,q_2,\cdots)$ 是平稳分布, 而由于定理 1 是唯一平稳分布. 从而 $\mathbb{Q}=\mathbb{Q}_s$.

最后假设 $2\leqslant N<\infty$ 或 $N=\infty$. 如果 $2\leqslant N<\infty$, 则以 $C^1,\cdots,C^N$ 表示正常返子类; 如果 $N=\infty$, 则以 $C^1,C^2,\cdots$ 表示正常返子类.

设数组 $\mathbb{Q}^k = (q_1^k, q_2^k, \cdots)$ 是类 C^k 中平稳分布, 由公式 (对照 (5) 式和 (6) 式)

$$q_j^k = \begin{cases} \mu_j^{-1} > 0, & j \in C^k, \\ 0, & j \notin C^k \end{cases}$$

建立的平稳分布. 那么, 对于任意非负数 $a_1, a_2, \cdots : a_1 + a_2 + \cdots = 1$(若 $N < \infty$, 则 $a_{N+1} = \cdots = 0$), 数组 $\{a_1\mathbb{Q}^1, \cdots, a_N\mathbb{Q}^N, \cdots\}$ 显然是平稳分布. 因而, 若假设 $2 \leqslant N < \infty$, 则导致 "一切平稳分布的集合" 是*连续统*, 但这与所作假设 "平稳分布唯一" 矛盾.

因而, 所作证明表示, 只可能是 $N = 1$. 换句话说, *如果存在 (唯一) 平稳分布, 则链恰好有一个由正常返状态形成的不可约类*.

充分性 假如马尔可夫链有不可约正常返状态子类, 即有情形 $N = 1$, 则由前面的叙述可见, (由定理 1 的命题 (c)) 平稳分布存在而且唯一.

于是, 定理 2 完全地得证. □

4. 定理 3 的证明 实质上, 这一定理的证明全部必要的, 都包含在定理 2 及其证明的论述中.

充分性 如果利用定理 2 的证明中记号, 则由定理的条件, 有 $N = 1, C = E$ 和 $d(E) = 1$ (非周期性). 那么, 由定理 2 的证明中关于 " $N = 1$ 的情形" 的讨论, 可见 $\mathbb{Q} = (q_1, q_2, \cdots)$, 其中 $q_j = \mu_j^{-1}, j \in E$, 是平稳分布, 同时又是遍历分布, 因为所有 $\mu_j^{-1} < \infty, j \in E$.

因此, 遍历分布 $\Pi = (\pi_1, \pi_2, \cdots)(= \mathbb{Q})$ 的存在性得证.

必要性 如果存在遍历分布 $\Pi = (\pi_1, \pi_2, \cdots)$, 则根据定理 1, 存在并且唯一等同于 Π 的平稳分布 $\mathbb{Q}$.

由定理 2 的论断 (及其证明), 可见 $N = 0$ 和 $2 \leqslant N < \infty$ 不可能出现, 即只有 $N = 1$, 并且仅存在唯一由正常返类组成的不可约类 C. 只剩下证明 $C = E$ 和 $d(E) = 1$.

用反证法, 假设 $C \neq E$ 和 $d(C) = 1$, 则仍然由于在定理 2 的证明中关于 " $N = 1$ 的情形" 的讨论, 可见存在状态 $j \notin C$, 使对一切 $i \in E$, 由 $p_{ij}^{(n)} \to 0$. 然而, 这与对于一切 $i \in E, \pi_j = \lim\limits_n p_{ij}^{(n)} > 0$ 矛盾.

这样, 在 $d(C) = 1$ 的情形下, 有 $C = E$ 和 $d(E) = 1$ (非周期性).

最后, 如果 $C \neq E$ 和 $d(C) > 1$, 则仍然由于在定理 2 的证明中关于 " $N = 1$ 的情形" 的讨论, 可见存在平稳分布 $\mathbb{Q} = (q_1, q_2, \cdots)$, 其中某些 $q_j = 0$, 这与 " $\Pi = \mathbb{Q}$, 而 $\Pi = (\pi_1, \pi_2, \cdots)$ 是遍历分布且 (根据定义) 所有 $\pi_j > 0, j \in E$" 矛盾. □

5. 定理 3 的推广 根据平稳 (不变) 分布 $\mathbb{Q} = (q_1, q_2, \cdots)$ 本身的定义, 是满足条件

$$q_j \geqslant 0, \quad j \in E = \{1, 2, \cdots\}, \quad \sum_{j=1}^{\infty} q_j = 1 \tag{10}$$

的一组数, 并且服从方程:

$$q_j = \sum_{i=1}^{\infty} q_i p_{ij}, \quad j \in E. \tag{11}$$

换一种提法, 可以认为, 平稳分布 $\mathbb{Q} = (q_1, q_2, \cdots)$ 是方程组

$$x_j = \sum_{i=1}^{\infty} x_i p_{ij}, \quad j \in E \tag{12}$$

的一个解, 其中方程组服从非负性条件和规范性条件:

$$x_j \geqslant 0, \quad j \in E \quad 和 \quad \sum_{j=1}^{\infty} x_j = 1.$$

如果满足定理 3 的条件, 则平稳解存在, 并且同时也是遍历的. 因此, 由定理 1 的命题 (c) 可见, 方程组 (12) 在序列类

$$(x_1, x_2, \cdots), \quad x_j \geqslant 0, \quad j \in E, \quad \sum_{j=1}^{\infty} x_j = 1$$

中有解并且唯一.

实际上, 由此可以得到更多结论. 具体地说, 这里定理 3 的条件也成立, 从而存在遍历分布 $\Pi = (\pi_1, \pi_2, \cdots)$.

在此条件下, 我们考虑在如下 (更广泛的) 序列类

$$(x_1, x_2, \cdots), \quad x_j \in \mathbb{R}, \quad j \in E, \quad \sum_{j=1}^{\infty} |x_j| < \infty, \quad \sum_{j=1}^{\infty} x_j = 1$$

中, 方程组 (12) 解的存在性问题. 我们证明在该类中, 解唯一并且遍历分布 Π 就是该唯一解.

事实上, 如果 $(x_1, x_2, \cdots)$ 是解, 则由于 $\sum\limits_{j=1}^{\infty} |x_j| < \infty$, 可以得到如下一系列等式:

$$\begin{aligned} x_j &= \sum_{i=1}^{\infty} x_i p_{ij} = \sum_{i=1}^{\infty} \left(\sum_{k=1}^{\infty} x_k p_{kj} \right) p_{ij} = \sum_{k=1}^{\infty} x_k \left(\sum_{i=1}^{\infty} p_{ki} p_{ij} \right) \\ &= \sum_{k=1}^{\infty} x_k p_{kj}^{(2)} = \cdots = \sum_{k=1}^{\infty} x_k p_{kj}^{(n)}, \end{aligned}$$

其中 $n \geqslant 1$. 当 $n \to \infty$ 时求极限, (根据控制收敛定理) 对于任意 $k \in E$, 由此得

$$x_j = \left(\sum_{k=1}^{\infty} x_k \right) \pi_j, \quad \pi_j = \lim_n p_{kj}^{(n)}.$$

由于根据假设 $\sum\limits_{k=1}^{\infty} |x_k| = 1$, 故 $x_j = \pi_j, j \in E$, 而这正是需要证明的.

6. 练习题

1. 对于转移概率矩阵为

$$\mathbb{P}=\begin{pmatrix}1/2 & 0 & 1/2 & 0\\ 0 & 0 & 0 & 1\\ 1/4 & 1/2 & 1/4 & 0\\ 0 & 1/2 & 1/2 & 0\end{pmatrix}$$

的马尔可夫链, 讨论其平稳分布, 极限分布, 遍历分布问题.

2. 设 $\mathbb{P}=(p_{ij})$ 是有限双随机矩阵 (即对于 $i=1,\cdots,m, \sum\limits_{j=1}^{m} p_{ij}=1$; 而对于 $i=1,\cdots,m, \sum\limits_{j=1}^{m} p_{ij}=1$). 证明, 对于相应的马尔可夫链, 向量 $\mathbb{Q}=(1/m,\cdots,1/m)$ 是平稳分布.

3. 设 $E=\{0,1\}$ 是马尔可夫链的状态空间, 而

$$\mathbb{P}=\begin{pmatrix}\alpha & 1-\alpha\\ 1-\beta & \beta\end{pmatrix},\quad 0<\alpha<1,\quad 0<\beta<1$$

是其转移概率矩阵. 讨论该马尔可夫链的, 极限分布, 遍历分布, 平稳分布问题.

§7. 有限马尔可夫链的极限分布、遍历分布和平稳分布

1. 有限链转移概率的渐近性质 根据 §5 定理 6, 任意不可约和非周期的、具有有限状态集的马尔可夫链, 都是正常返的. 这一事实使得有可能, 给 §6 的定理 3 以如下表述.(对照 §3 的问题, **A**, **B**, **C** 和 **D**).

定理 1 假设 $X=(X_n)_{n\geqslant 0}$ 是具有有限状态集 $E=\{1,2,\cdots,r\}$ 的马尔可夫链, 并且是不可约的和非周期的.

那么, 如下命题成立.

(a) 对于一切 $j\in E$, 存在不依赖于初始状态 $i\in E$ 的极限值 $\pi_j=\lim\limits_n p_{ij}^{(n)}$.

(b) 极限值 $\Pi=(\pi_1,\cdots,\pi_r)$ 是概率分布, 即 $\pi_j\geqslant 0, \sum\limits_{i=1}^{r}\pi_i=1, j\in E$.

(c) 此外, 对于一切 $j\in E$, 极限值 $\pi_j=\mu_j^{-1}>0$, 其中 μ_j 是首返状态 j 前度过的平均时间:

$$\mu_j=\sum_{n=1}^{\infty} n f_{jj}^{(n)}=\mathbf{E}_j\tau(j),$$

其中 $\tau(j)=\inf\{n\geqslant 1: X_n=j\}$. 从而, 数组 $\Pi=(\pi_1,\cdots,\pi_r)$ 是遍历分布.

(d) 平稳分布 $\mathbb{Q}=(q_1,\cdots,q_r)$ 存在、唯一并且等于 $\Pi=(\pi_1,\cdots,\pi_r)$.

2. 不可约性和非周期性对有限链的意义　作为定理 1 的补充, 我们引进如下定理, 其作用体现在 "不可约性" 和 "非周期性" 上.

定理 2　假设 $X = (X_n)_{n\geqslant 0}$ 是具有有限状态集 $E = \{1, 2, \cdots, r\}$ 的马尔可夫链, 并且是不可约的和非周期的.

那么, 如下命题成立.

(a) 链是不可约的和非周期的 $(d = 1)$;

(b) 链是不可约的, 非周期的 $(d = 1)$, 正常返的;

(c) 链是遍历的;

(d) 存在 n_0, 使对于一切 $n \geqslant n_0$, 有

$$\min_{i,j\in E} p_{ij}^{(n)} > 0.$$

证明　蕴含关系 (d)⇒(c), 在第一章 §12 定理 1 中已经证明. 相反的蕴含关系 (c)⇒(d) 显然. 蕴含关系(a)⇒(b), 由 §5 定理 6 可见; 而蕴含关系 (b)⇒(a) 显然. 最后, 命题 (b) 与 (c) 等价, 包含在 §6 定理 3 中. □

§8. 作为马尔可夫链的简单随机游动

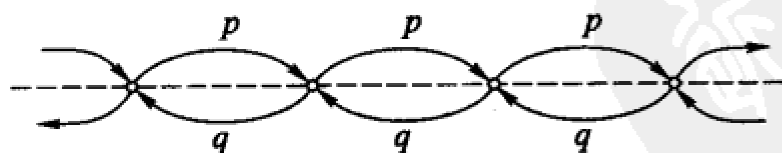
1. 简单随机游动 · 波利亚 (G. Polia) 定理　$d-$ 维简单随机游动指, 描绘 "质点" 在格子的结点 $\mathbb{Z}^d = \{0, \pm 1, \pm 2, \cdots\}^d$ 上运动的、齐次马尔可夫链, 这是该 "质点" 可能以某一概率逗留在每一状态, 也可能以一定概率转移到相邻的主题之一.

例 1　设 $d = 1$, 而链的状态集为 $E = \mathbb{Z} = \{0, \pm 1, \pm 2, \cdots\}$, 其转移概率矩阵有如下形式:

$$p_{ij} = \begin{cases} p, & j = i + 1, \\ q, & j = i - 1, \\ 0, & \text{其他}, \end{cases}$$

其中 $p + q = 1$.

该矩阵对应于图



它直观上描绘该链可能的转移.

如果 $p = 0$, 则指点肯定向左移; 如果 $p = 1$, 则指点肯定向右移.

这些 "确定性" 情形没有什么意义, 况且这时的一切状态都是非本质的. 因此, 我们将假设 $0 < p < 1$.

在这样的条件下, 链由一个本质互通状态类组成. 换句话说, 在假设 $0 < p < 1$ 下, 链是不可约的 (见 §4).

对于任意 $j \in E$, 由二项分布的公式 (第一章 §2), 可见

$$p_{jj}^{(2n)} = \mathrm{C}_{2n}^{n}(pq)^n = \frac{(2n)!}{(n!)^2}(pq)^n. \tag{1}$$

根据司特林公式 (第一章 §2 的 (6) 式; 亦见练习题 1):

$$n! \sim \sqrt{2\pi n}n^n \mathrm{e}^{-n}.$$

从而, 由 (1) 式可见

$$p_{jj}^{(2n)} \sim \frac{(4pq)^n}{\sqrt{\pi n}}, \tag{2}$$

因此

$$\sum_{n=1}^{\infty} p_{jj}^{(2n)} = \infty, \quad 若\ p = q. \tag{3}$$

$$\sum_{n=1}^{\infty} p_{jj}^{(2n)} < \infty, \quad 若\ p \neq q. \tag{4}$$

由这些公式和 §5 的定理 1, 得到如下结果:

在集合 $E = \mathbb{Z} = \{0, \pm1, \pm2, \cdots\}$ 上的简单一维随机游动, 若 $p = q = 1/2$, 则是常返的和对称的; 而若 $p \neq q$, 则是非常返的.

在第一章 §10 曾经证明对于 $p = q = 1/2$ 的情形, 当 n 充分大时

$$f_{jj}^{(2n)} \sim \frac{1}{2\sqrt{\pi}n^{3/2}}. \tag{5}$$

因此

$$\mu_j = \sum_{n=1}^{\infty}(2n)f_{jj}^{(2n)} = \infty, \quad j \in E. \tag{6}$$

从而, 这时一切状态是*常返的和零的*. 因此根据 §5 的定理 5, 在一切 $0 < p < 1$ 的情形下, 对于任意 i 和 j, 当 $n \to \infty$ 时 $p_{ij}^{(n)} \to 0$. 由此可见 (§6 的定理 1), 极限分布, 以及平稳分布和遍历分布*都不存在*.

例 2 设 $d = 2$. 考虑对称的情形 (对应于例 1 中 $p = q = 1/2$ 的情形): 质点向左、向右的、向上或向下移动概率都等于1/4.

为确定计, 固定零状态 $\mathbf{0} = (0, 0)$. 假设开始处于零状态, 并讨论 "质点" *常返*或者*非常返*零状态的问题.

为此, 我们讨论游动质点如下的那些 "轨道": 质点向右移动 i 步、向左 i 步、向上 j 步和向下 j 步. 假如 $2i + 2j = 2n$, 则说明 "质点" 从零点出发, 经过 $2n$ 步必然又返回该状态. 显然, 经奇数步 "质点" 不可能返回零状态.

由此可见, 从状态 $\mathbf{0}$ 仍然返回状态 $\mathbf{0}$ 的转移概率决定于下面的公式:

$$p_{\mathbf{00}}^{(2n+1)} = 0, \quad n = 0, 1, 2, \cdots$$

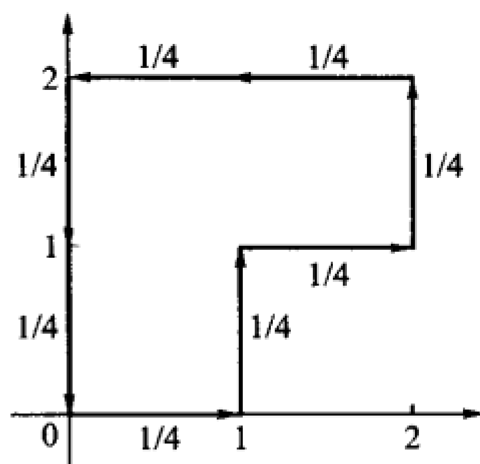


图 41　平面上的游动

且 (由全概率公式)

$$p_{\mathbf{00}}^{(2n)} = \sum_{(i,j):i+j=n} \frac{(2n)!}{(i!)^2(j!)^2}\left(\frac{1}{4}\right)^{2n}, \quad n = 1, 2, \cdots \tag{7}$$

(亦见第一章 §2 第 2 小节).

在 (7) 式中的和号下分式的分子与分母同乘以 $(n!)^2$，得

$$p_{\mathbf{00}}^{(2n)} = \left(\frac{1}{4}\right)^{2n} \mathrm{C}_{2n}^{n} \sum_{i=0}^{n} \mathrm{C}_{n}^{i}\mathrm{C}_{n}^{n-i} = \left(\frac{1}{4}\right)^{2n} (\mathrm{C}_{2n}^{n})^2, \tag{8}$$

其中用到公式:

$$\sum_{i=0}^{n} \mathrm{C}_{n}^{i}\mathrm{C}_{n}^{n-i} = \mathrm{C}_{2n}^{n}$$

(第一章 §2 练习题 4).

根据司特林公式, 由 (8) 式可得 $p_{\mathbf{00}}^{(2n)} \sim \dfrac{1}{\pi n}$, 因而

$$\sum_{n=0}^{\infty} p_{\mathbf{00}}^{(2n)} = \infty. \tag{9}$$

根据对称性, 类似的论断, 当然不仅对零状态成立, 而且对于任何状态 (i, j) 也成立.

像 $d = 1$ 的情形一样, 由 §5 定理 1 的公式 (9), 得到如下结果:

在集合 $E = \mathbb{Z}^2 = \{0, \pm1, \pm2, \cdots\}^2$ 上的简单双对称随机游动是常返的.

例 3　结果表明, 当 $d \geqslant 3$ 时, 在状态 $E = \mathbb{Z}^d = \{0, \pm1, \pm2, \cdots\}^d$ 上对称随机游动, 非常明显的不同于上面讨论的 $d = 1$ 和 $d = 2$ 的情形.

具体地说,

在集合 $E = \mathbb{Z}^d = \{0, \pm1, \pm2, \cdots\}^d$ 上的简单 $d-$ 维对称随机游动, 对于任意 $d \geqslant 3$ 是非常返的.

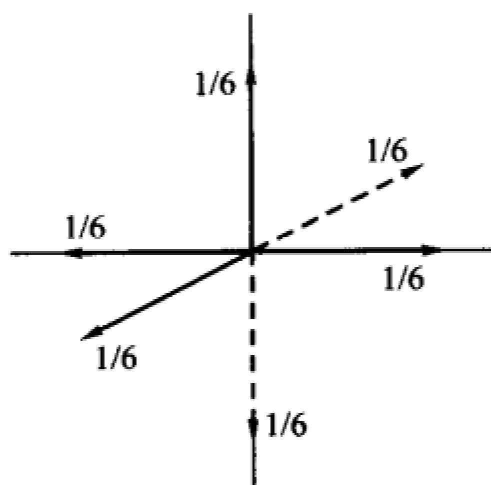
这一事实的证明基于如下考虑: 对于 $d \geqslant 3$, 当 $n \to \infty$ 时, 概率 $p_{jj}^{(2n)}$ 有如下渐近式:

$$p_{ij}^{(2n)} \sim \frac{c(d)}{n^{d/2}}, \tag{10}$$

其中 $c(d)$ 是某一依赖于维数 d 的正常数.

我们现在证明 $d=3$ 的情形, 而 $d>3$ 的情形的证明留做练习题.

由于随机游动对称性的假设, 可见 "质点" 以概率 1/6 沿 (空间 3 条) 坐标轴的六个方向之一移动一步:



假设 "质点" 从点 $\mathbf{0}$=(0,0,0) 出发. 那么, 像 $d=2$ 的情形, 由多项分布公式 (第一章 §2), 有

$$\begin{aligned} p_{\mathbf{00}}^{(2n)} &= \sum_{(i,j):0\leqslant i+j\leqslant n} \frac{(2n)!}{(i!)^2(j!)^2[(n-i-j)!]^2}\left(\frac{1}{6}\right)^{2n} \\ &= 2^{-2n}\mathrm{C}_{2n}^{n} \sum_{(i,j):0\leqslant i+j\leqslant n} \left[\frac{n!}{i!j!(n-i-j)!}\right]^2 \left(\frac{1}{3}\right)^{2n} \\ &\leqslant \mathrm{C}_n 2^{-2n}\mathrm{C}_{2n}^{n} 3^{-n} \sum_{(i,j):0\leqslant i+j\leqslant n} \frac{n!}{i!j!(n-i-j)!}\left(\frac{1}{3}\right)^{2n} \\ &= \mathrm{C}_n 2^{-2n}\mathrm{C}_{2n}^{n} 3^{-n}, \end{aligned} \tag{11}$$

其中

$$\mathrm{C}_n = \max_{(i,j):0\leqslant i+j\leqslant n} \left[\frac{n!}{i!j!(n-i-j)!}\right], \tag{12}$$

并且利用了明显的公式:

$$\sum_{(i,j):0\leqslant i+j\leqslant n} \frac{n!}{i!j!(n-i-j)!}\left(\frac{1}{3}\right)^{2n} = 1.$$

下面将要证明

$$\mathrm{C}_n \sim \frac{n!}{[(n/3)!]^3}. \tag{13}$$

运用司特林公式, 由 (13) 式, 可得

$$\mathrm{C}_n 2^{-2n}\mathrm{C}_{2n}^n 3^{-n} \sim \frac{3\sqrt{3}}{2\pi^{3/2}n^{3/2}}. \tag{14}$$

因而, 由 (11) 式, 可得

$$\sum_{n=1}^{\infty} p_{\mathbf{00}}^{(2n)} < \infty \tag{15}$$

于是, 根据 §5 的定理 1, 状态 $\mathbf{0}$=(0,0,0) 是*非常返的*. 由于对称性, 对于 $E=\mathbb{Z}^3$ 中的任何其他状态, 有类似的结果.

只剩下证明 (13) 式.

设

$$m_n(i,j) = \frac{n!}{i!j!(n-i-j)!},$$

而 $i_0 = i_0(n), j_0 = j_0(n)$ 的值满足

$$\max_{(i,j):0\leqslant i+j\leqslant n} m_n(i,j) = m_n(i_0,j_0).$$

取 4 个点 $(i_0-1,j_0),(i_0+1,j_0),(i_0,j_0-1),(i_0,j_0+1)$. 由于相应的值 $m_n(i_0-1,j_0),m_n(i_0+1,j_0),m_n(i_0,j_0-1)$ 和 (i_0,j_0+1) 不大于 $m_n(i_0,j_0)$, 可得 4 个不等式:

$$n-i_0-1 \leqslant 2j_0 \leqslant n-i_0+1,$$
$$n-j_0-1 \leqslant 2i_0 \leqslant n-j_0+1.$$

由这些不等式, 可得

$$i_0(n) \sim \frac{n}{3}, \quad j_0(n) \sim \frac{n}{3},$$

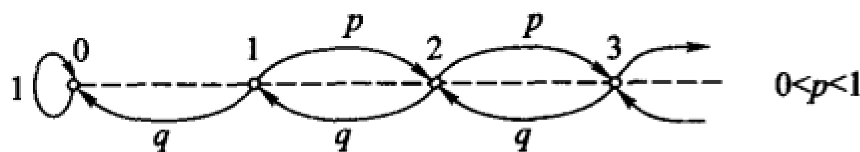
由此得所要证明得 (13) 式.

归纳所分析 $d=1,2,3$ 的情形, 得如下波利亚 (G. Polia) 的结果.

定理 *当 $d=1$ 或 $d=2$ 时, 在状态集 $E=\mathbb{Z}^d=\{0,\pm1,\pm2,\cdots\}^d$ 上的简单对称随机游动是常返的, 而当 $d=3$ (及 $d\geqslant3$) 时是非常返的.*

2. 简单随机游动的例 ($E\subset\mathbb{Z}^d, d=1$) 上面的例子讲的是, 在 "整个" 空间 $\mathbb{Z}^d$ 中的简单随机游动. 在这一小节, 将要讨论的简单随机游动, 其相空间 E *严格*包含在 $\mathbb{Z}^d$ 之内. 这里, 我们讲局限于 $d=1$ 的情形.

例 4 考虑简单随机游动, 其相空间 $E=\{0,1,2,\cdots\}$, 其中状态 0 是*吸收的*, 而其转移概率如下图所示:



这里, 状态 0 是正常返状态, 并且构成唯一不可约子类. (其余状态都是非常返的.) 根据 §6 的定理 2, 存在并且唯一平稳分布 $\mathbb{Q} = (q_0, q_1, \cdots), q_0 = 1, q_i = 0(i = 1, 2, \cdots)$.

所考虑的随机游动直观上提供了一个例子: 对于某 i 和 j, 极限 $\lim\limits_n p_{ij}^{(n)}$ 存在, 但是依赖于初始状态, 同时说明在此随机游动的例子中不存在遍历分布..

显然, $p_{00}^{(n)} = 1$, 而对于 $j = 1, 2, \cdots, p_{0j}^{(n)} = 0$. 经简单运算, 可见对于一切 $i, j = 1, 2, \cdots, p_{ij}^{(n)} \to 0$.

现在证明对于一切 $i = 1, 2, \cdots$, 极限 $\alpha(i) = \lim\limits_n p_{i0}^{(n)}$ 存在, 并且对其有如下公式:

$$\alpha(i) = \begin{cases} (q/p)^i, & p > q, \\ 1, & p \leqslant q. \end{cases} \tag{16}$$

由该式可见, 对于 $p > q$ 的情形 ("有向右运动的倾向"), 由状态 $i(i = 1, 2, \cdots)$ 向状态 0 转移的极限概率 $\alpha(i) = \lim\limits_n p_{i0}^{(n)}$, 事实上依赖于 i, 并且随着 i 的增长以几何的速度减小.

为证明 (16) 式, 首先注意到, 由于状态 0 是吸收状态, 故 $p_{i0}^{(n)} = \sum\limits_{k \leqslant n} f_{i0}^{(k)}$, 从而极限 $\alpha(i) = \lim\limits_n p_{i0}^{(n)}$ 存在, 并且等于 f_{i0}, 即概率 $\alpha(i)$ 是 "质点" 由状态 i 出发, 迟早到达状态 "零" 的概率. 对于这些概率, 在第一章 §12 (亦见第七章 §2), 当 $\alpha(i) = 1$ 时导出了递推公式:

$$\alpha(i) = p\alpha(i+1) + q\alpha(i-1). \tag{17}$$

该方程的通解为

$$\alpha(i) = a + b(q/p)^i, \tag{18}$$

而由条件 $\alpha(0) = 1$, 可见常数 a 和 b 满足条件 $a + b = 1$.

如果假设 $q > p$, 则因为 $\alpha(i)$ 有界, 立即得 $b = 0$, 所以 $\alpha(i) = 1$. 这一结论十分清楚, 因为当 $q > p$ 时, "质点" 有沿到达零状态运动的倾向.

假如 $p > q$, 则情况相反: 有向右移动的倾向, 且自然想到

$$\alpha(i) \to 0, \quad i \to \infty, \tag{19}$$

即 $\alpha = 0$, 而

$$\alpha(i) = (q/p)^i. \tag{20}$$

为证明该等式, 我们并不先证明 (19) 式, 而是通过其他途径.

除在点 0 的吸收屏幕外, 我们在引进在点 N 的吸收屏幕. 以 $\alpha_N(i)$ 表示, 从点 i 出发的 "质点", 到达状态 0 早于到达状态 N 的概率. 概率 $\alpha_N(i)$ 满足边界条件为

$$\alpha_N(0) = 1, \quad \alpha_N(N) = 0$$

的方程 (17), 而且在第一章 §9 已经证明

$$\alpha_N(i) = \frac{\left(\frac{q}{p}\right)^i - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N}, \quad 0 \leqslant i \leqslant N. \tag{21}$$

由此可见 $\lim\limits_N \alpha_N(i) = (q/p)^i$, 从而为证明 (20) 式的结果, 只需证明

$$\alpha(i) = \lim_N \alpha_N(i). \tag{22}$$

这在直观上很清楚. 严格的证明有如下一些途径.

假设 "质点" 从固定 i 状态出发. 那么,

$$\alpha(i) = \mathbf{P}_i(A), \tag{23}$$

其中 A 是事件 "存在 N, 使由 i 点出发的 '质点' 到达状态 0 早于到达状态 N ". 如果

$$A = \{\text{“质点” 到达状态 0 早于状态} N\},$$

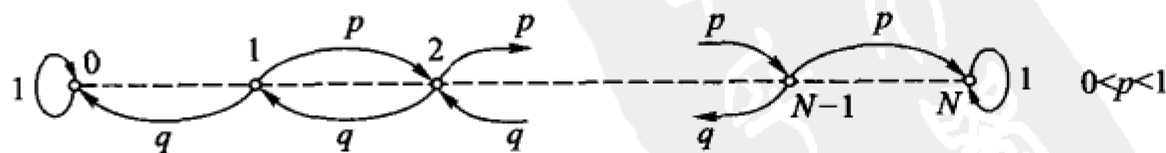
则 $A = \bigcup\limits_{N=i+1}^{\infty} A_N$. 显然, $A_N \subseteq A_{N+1}$, 且

$$\mathbf{P}_i\left(\bigcup_{N=i+1}^{\infty} A_N\right) = \lim_{N\to\infty} \mathbf{P}_i(A_N). \tag{24}$$

因为 $\alpha_N(i) = \mathbf{P}_i(A_N)$, 所以由 (23) 式和 (24) 式, 立即得 (22) 式.

这样, 如果 $p > q$, 则极限值 $\lim\limits_n p_{i0}^{(n)}$ 依赖于 i. 如果 $p \leqslant q$, 则对于任意 i, 极限值 $\lim\limits_n p_{i0}^{(n)} = 1$, 且 $\lim\limits_n p_{ij}^{(n)} = 0, j \geqslant 1$. 于是, 在这种情形下存在不依赖于 i 的极限分布 $\Pi = (\pi_0, \pi_1, \cdots)$, 其中 $\pi_j = \lim\limits_n p_{ij}^{(n)}$. 这时 $\Pi = (1, 0, 0, \cdots)$.

例 5 考虑简单随机游动, 其相空间 $E = \{0, 1, \cdots, N\}$, 其中 "边界" 状态 0 和 N 是吸收的:



这里, 存在两个不可约的正常返类: $\{0\}$ 和 $\{N\}$, 而所有其他状态类 $1, 2, \cdots, N-1$ 都是非常返的. 由 §6 定理 2 的证明可见, 存在平稳分布 $\mathbb{Q} = (q_0, q_1, \cdots, q_N)$ 的连续统, 且其平稳分布都具有如下形式: $q_1 = \cdots = q_{N-1} = 0, q_0 = a, q_N = b, a \geqslant 0, b \geqslant 0, a + b = 1$.

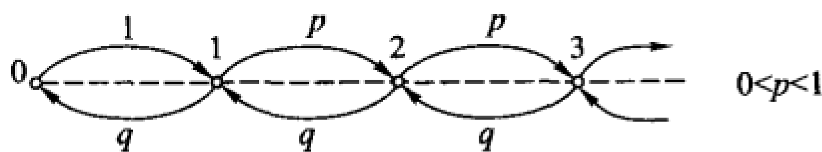
根据第一章 §9 第 2 小节, 有

$$\lim_n p_{i0}^{(n)} = \begin{cases} \dfrac{\left(\dfrac{q}{p}\right)^i - \left(\dfrac{q}{p}\right)^N}{1 - \left(\dfrac{q}{p}\right)^N}, & p \neq q, \\ 1 - \dfrac{i}{N}, & p = q = \dfrac{1}{2}, \end{cases} \tag{25}$$

$$\lim_n p_{iN}^{(n)} = 1 - \lim_n p_{i0}^{(n)} \text{ 和 } \lim_n p_{ij}^{(n)} = 0, 1 \leqslant j \leqslant N-1.$$

需要强调, 这里及前面例子中, 极限值 $\lim_n p_{ij}^{(n)}$ 依赖于初始状态.

例 6 考虑简单随机游动, 其相空间 $E = \{0, 1, \cdots\}$, 且在状态 0是*反射壁*:



所考虑的链本质上依赖于 p 和 q.

如果 $p > q$, 则游动的 "质点" 有向右移动的倾向. 而在 "零" 状态存在*反射壁*的情形, 与例 4 中游动相比仅仅在于, 对于例 4 的情形在 "零" 状态可以越过 "壁障". 这是全部状态都是非常返的; 对于一切 $i, j \in E, p_{ij}^{(n)} \to 0, n \to \infty$; 平稳分布和遍历分布都不存在.

如果 $p < q$, 则游动的质点有向左移动的倾向. 这时, 链是常返的. 当 $p = q$ 时链也是常返的.

现在给出平稳分布 $\mathbb{Q} = (q_0, q_1, \cdots)$ 应满足的方程组 (对照 §6 的 (12) 式):

$$\begin{aligned} q_0 &= q_1 q, \\ q_1 &= q_0 + q_2 q, \\ q_2 &= q_1 p + q_3 q, \\ &\cdots\cdots\cdots\cdots \end{aligned}$$

由此, 有

$$\begin{aligned} q_1 &= q(q_1 + q_2), \\ q_2 &= q(q_2 + q_3), \\ &\cdots\cdots\cdots\cdots \end{aligned}$$

于是

$$q_j = \left(\frac{p}{q}\right) q_{j-1}, \quad j = 2, 3, \cdots$$

假如 $p=q$, 则 $q_1=q_2=\cdots$, 从而方程组没有满足条件

$$\sum_{j=1}^{\infty} q_j = 1, \quad q_0 = q_1 q$$

的非负解.

这意味着, 当 $p=q=1/2$ 时不存在平稳解. 这时链的全部状态都是常返的.

最后, 设 $p<q$. 由条件 $\sum\limits_{j=0}^{\infty} q_j = 1$, 可见

$$q_1 \left[q + 1 + \frac{p}{q} + \left(\frac{p}{q}\right)^2 + \cdots \right] = 1.$$

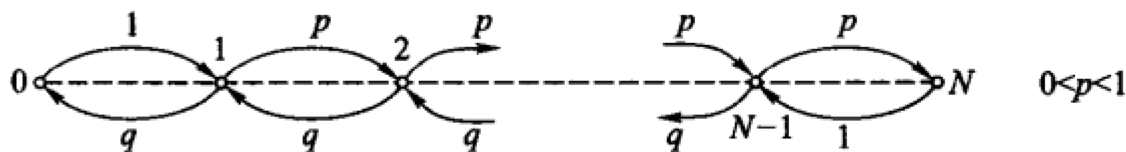
因此

$$q_1 = \frac{q-p}{2q}, \quad q_0 = q_1 q = \frac{q-p}{2},$$

而

$$q_j = \frac{q-p}{2q} \left(\frac{p}{q}\right)^{j-1}, \quad j \geqslant 2.$$

例 7 考虑简单随机游动, 相空间为 $E=\{0,1,\cdots,N\}$, 而状态 0 和 N 是反射壁:



这里, 链的状态是一个不可约类. 状态是周期为 $d=2$ 的正常返的. 根据 §6 的定理 2, 该链有唯一平稳分布 $\mathbb{Q}=(q_0,q_1,\cdots,q_N)$. 在条件

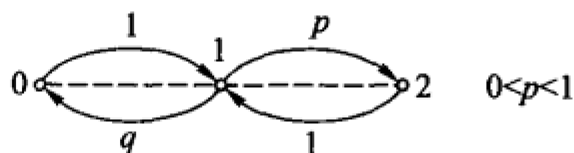
$$\sum_{i=0}^{\infty} q_i = 1, \quad q_j \geqslant 0, \quad j \in E,$$

下, 解方程组 $q_j = \sum\limits_{i=0}^{N} q_i p_{ij}$, 得

$$q_j = \frac{\left(\dfrac{p}{q}\right)^{j-1}}{1 + \sum\limits_{i=1}^{N-1} \left(\dfrac{p}{q}\right)^{i-1}}, \quad 1 \leqslant j \leqslant N-1, \tag{26}$$

而 $q_0 = q_1 q, q_N = q_{N-1} q$.

由 §6 的定理 3, 以及所考虑链的周期为 $d=2$, 可见遍历分布不存在. 可以直接得出结论, 这里无遍历分布. 例如, 对于 $N=2$



那么, 易见 $p_{11}^{(2n)}=1$, 而 $p_{11}^{(2n+1)}=0$. 因而 $\lim\limits_{n} p_{11}^{(2n)}$ 不存在. 与此同时, 由 (26) 式可见, 平稳分布:

$$\mathbb{Q}=(q_0,q_1,q_2),$$

并且具有如下形式:

$$q_0=\frac{1}{2}q,\quad q_1=\frac{1}{2},\quad q_2=\frac{1}{2}p.$$

3. 利用简单随机游动描绘现实物理过程的示例 书中叙述的资料表明, 简单随机游动是一个经典模型. 基于这一模型, "精练了" 概率的思想体系, 使得概率的技术更加完美, 发现了许多概率 - 统计的规律性. 例如, 考虑只有两个可能值的伯努利随机变量 $\xi_1,\xi_2,\cdots$ 之和 $X_n=\xi_1+\cdots+\xi_n$, 从而, $X=(X_n)_{n\geqslant 1}$ 是简单随机游动 (也是马尔可夫链). 基于 $X=(X_n)_{n\geqslant 1}$, 发现了许多规律性, 诸如, 大数定律 (第一章 §5), 棣莫弗 - 拉普拉斯定理 (第一章 §6), 反正弦定律 (第一章 §10), 等等.

在这一小节, 我们考虑两个离散扩散模型, 它们是 "利用简单随机游动, 如何反映现实的物理过程" 的很好示例.

A. 埃伦弗斯特 (P. Ehrenfest, T. Ehrenfest) 模型

像例 7 一样, 我们将考虑简单随机游动, 假设其相空间为 $E=\{0,1,\cdots,N\}$, 且在状态 0 和 N 各有一个反射壁.

在状态 0 和 N 的转移概率为 $p_{01}=1$ 和 $p_{N,N-1}=1$. 在其余状态 $i=1,\cdots,N-1$, "质点" 只可能以概率

$$p_{ij}=\begin{cases}1-\dfrac{i}{N}, & j=i+1,\\ \dfrac{i}{N}, & j=i-1\end{cases}\tag{27}$$

向左或向右移动一步.

在 1907 年 P. 埃伦弗斯特和 T. 埃伦弗斯特 [124], 在研究统计力学的一种模型时, 得到了具有上述转移概率的马尔可夫链. 这一统计力学中描绘气体分子的模型是: 仪器有两个 (封闭的) 箱子 A 和 B, 由薄膜连接在一起, 气体分子由一个箱子 (A 或 B), 通过薄膜的 (微) 小孔移动到另一个箱子 (B 或 A).

假设在所观察的两个箱子中, 分子的总数等于 N, 且每一步从一个箱子向另一个箱子的移动是以如下的方式积极地进行的: 以概率 $1/N$ 随机地选一个分子, 使它转移到另一个箱子; 并且每一步欲转移分子的选取, 与过去的 "历史" 无关.

设 X_n 是这样分子的个数: (例如) 分子一 "开始" 于时刻 n 在某一个箱子 A 中, 为描绘分子的移动的机制有马尔可夫性 (练习题 2):

$$\mathbf{P}\{X_{n+1}=j|X_0=i_0,X_1=i_1,\cdots,X_{n-1}=i_{n-1},X_n=i\}$$
$$=\mathbf{P}\{X_{n+1}=j|X_n=i\}, \tag{28}$$

此外,

$$\mathbf{P}\{X_{n+1}=j|X_n=i\}=p_{ij}, \tag{29}$$

其中 p_{ij} 由 (27) 式决定.

对于该模型存在由如下二项公式

$$q_j=\mathrm{C}_N^j\left(\frac{1}{2}\right)^N, \quad j=0,1,\cdots,N \tag{30}$$

决定的 (练习题3) 平稳分布 $\mathbb{Q}=(q_0,q_1,\cdots,q_N)$. 这里所有考虑的链是常返的 (练习题 4).

需要指出, 例如当 N 为偶数时. 概率 $q_j(j=0,1,\cdots,N)$ 在 "中间" 值 $N/2$ 上达到最大值, 这恰好对应于最可能的 "平衡" 状态, 即这时两个箱子中的分子数相同.

可以想到, 这样建立的随时间的 "平衡", 带有 (由上面引进的分布 $\mathbb{Q}$ 描绘的) 概率 – 统计特点.

我们还要指出, 在直观上分子的数量随时间 "稳定的" 可能性十分明显: 状态 i 离 "中心" 值越远, (由于 (27) 式) 向该值方向运动的概率就越大.

B. D. 伯努利 – 拉普拉斯模型

考虑在一定意义上与埃伦弗斯特模型类似的模型. 该模型是 D. 伯努利 (1769 年) 提出的, 后来拉普拉斯 (1812 年) 在描绘不可压缩液体 (而不是质点) 过程时提出来的.

更精确地说, 有两个样品箱 A 和 B, 含有 $2N$ 个质点, 其中N 个质点是 "白色" 的, 而另外 N 个质点是 "黑色" 的.

称 "系统" 处于状态 i, 其中 $i\in E=\{0,1,\cdots,N\}$, 如果在箱 A 中恰好有 i 个 "白色" 质点和 $N-i$ 个 "黑色" 质点. 假设 "不相容性" 是指, 对于所观测的状态 i, 在箱 B 中有 $N-i$ 个 "白色" 质点, 有 i 个 "黑色" 质点. 每箱中质点的总数保持为常数, 并且等于 N.

在每一步 n, 从每箱中随机地 (即以概率 $1/N$) 选一个质点, 并且这些点改变位置. 假设这一 (随机) 从箱中选择质点的过程是独立进行的, 并且以后的选择按同样的模式进行, 而且与以往的阶段无关.

以 X_n 表示 A 箱中 "白" 色质点的个数. 那么, 质点变换的机制导致马尔可夫性 (28) 的成立, 这是 (29) 式中转移概率 p_{ij} 决定于如下表达式 (练习题 5):

$$p_{ij} = \begin{cases} \left(\frac{i}{N}\right)^2, & j = i-1, \\ \left(1-\frac{i}{N}\right)^2, & j = i+1, \\ 2\frac{i}{N}\left(1-\frac{i}{N}\right), & j = i, \end{cases} \tag{31}$$

其中 $p_{ij} = 0$, 如果 $|i-j| > 1, i = 0, 1, \cdots, N$.

像埃伦弗斯特模型一样, 这里的所有状态也都是常返的. 平稳分布 $\mathbb{Q} = (q_0, q_1, \cdots, q_N)$ 存在而且唯一, 并且由如下公式表示 (练习题 5):

$$q_j = \frac{(\mathrm{C}_N^j)^2}{(\mathrm{C}_{2N}^N)^2}, \quad j = 0, 1, \cdots, N. \tag{32}$$

4. 现实物理过程的示例 (更加复杂的情形) 在这一章的开始曾经指出, 这一章的基本内容是, "(随着 n 的增长) 无后效系统的渐近性质", 上一节内容表明, 研究的是, 具有可数状态集 $E = \{0, 1, \cdots, N, \cdots\}$ 的马尔可夫链的, 转移概率 p_{ij} 当 n 充分大时的性质, 包括简单随机游动, 转移只可能发生在相邻状态之间的情形.

很重要的是, 对于具有更加复杂状态空间的马尔可夫链研究类似的问题. 关于这类内容, 例如可以见 [75], [117].

5. 关于术语 "离散扩散模型" 的说明 上面讨论的两个模型 (埃伦弗斯特模型和伯努利 – 拉普拉斯模型) 曾称之为*离散扩散模型*.

现在关于这一名称做一些说明, 为此考虑 $\mathbb{R}$ 中随机游动的极限性质. 设 $S_n = \xi_1 + \cdots + \xi_n, n \geqslant 1, S_0 = 0$, 其中 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量序列, 且 $\mathbf{E}\xi_i = 0$, $\mathbf{D}\xi_i = 1$. 设 $X_0^n = 0$, 而

$$X_t^n = \frac{S_{[nt]}}{\sqrt{n}} \quad \left(= \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} \xi_k\right), \quad 0 < t \leqslant 1.$$

显然, 序列 $(0, X_{1/n}^N, X_{2/n}^N, \cdots, X_1^N)$ 可以视为在时刻 $\Delta, 2\Delta, \cdots, 1$, 其中 $\Delta = 1/n$, 跃度的量级为 $\sqrt{\Delta}(\Delta X_{k\Delta}^n \equiv X_{k\Delta}^n - X_{(k-1)\Delta}^n = \xi_k\sqrt{\Delta})$ 的简单随机游动.

像在第七章 §8 注 4 已经指出的那样, 这一随机游动 $X^n = (X_t^n)_{0 \leqslant t \leqslant 1}$ 的全部有限维分布, 弱收敛于维纳过程 (布朗运动) $W = (W_t)_{0 \leqslant t \leqslant 1}$; 此外, 在第七章 §8 注 4 中还指出, 函数收敛性也成立, 即过程 X^n 之分布向过程 W 之分布的弱收敛 (收敛的意义, 仍然是经验过程向布朗桥的收敛性; 见第三章 §13 的第 4 小节). 维纳过程是*扩散过程*典型的 (并且是基本的) 例子 (见 [69, V. II], [21], [131]). 这一事实恰好说

明, X^n 类型的过程, 以及出现在埃伦弗斯特模型和伯努利 – 拉普拉斯模型中的, 过程为什么称为离散扩散模型.

6. 练习题

1. 利用如下概率的思想 ([106], 题 27.18), 证明斯特林公式: $n! \sim \sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}$. 设 $S_n = X_1 + \cdots + X_n (n \geqslant 1)$, 其中 $X_1, X_2, \cdots$ 是独立且服从参数 $\lambda = 1$ 的泊松分布的随机变量. 依次证明:

(a) $$\mathbf{E}\left(\frac{S_n - n}{\sqrt{n}}\right)^- = e^{-n}\sum_{k=0}^{n}\left(\frac{n-k}{\sqrt{n}}\right)\frac{n^k}{k!} = \frac{n^{n+\frac{1}{2}}}{n!}e^{-n};$$

(b) $$\text{Law}\left[\left(\frac{S_n - n}{\sqrt{n}}\right)^-\right] \to \text{Law}\left[N^-\right],$$

其中 N 是正态分布的随机变量;

(c) $$\mathbf{E}\left(\frac{S_n - n}{\sqrt{n}}\right)^- \to \mathbf{E}N^- = \frac{1}{\sqrt{2\pi}};$$

(d) $$n! \sim \sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}.$$

2. 证明马尔可夫性 (28) 式.

3. 证明 (30) 式.

4. 证明埃伦弗斯特模型中马尔可夫链的所有状态都是常返的.

5. 验证 (31) 式和 (32) 式的正确性.

§9. 马尔可夫链的最优停时问题

1. 这一节的基本内容 下面讨论的内容, 与第七章 §13 紧密衔接, 这部分阐述任意随机序列的、最优停时问题求解的 "鞅" 方法. 这一节的基本内容, 涉及由马尔可夫链之状态的函数产生的随机序列, 可以赋予第七章 §13 的结果简单而直观的形式和解释.

2. 一步转移算子 假设 $X = (X_n, \mathscr{F}_n, \mathbf{P}_x)$ 是具有相空间 $(E, \mathscr{E})$ 的、离散时间齐次马尔可夫链.

亦假设 $(\Omega, \mathscr{F})$ 是坐标空间 (见 §1 的第 6 小节), 而 $X_n = X_n(\omega), n \geqslant 0$, 是定义在 $(\Omega, \mathscr{F})$ 上的随机变量, 并且 $X_n = X_n(\omega)$ 本身是以坐标的形式给定的: 如果 $\omega = (x_0, x_1, \cdots) \in \Omega$, 则 $X_n(\omega) = x_n$; 这里 $\mathscr{F}$ 理解为 $\sigma-$ 代数: $\mathscr{F} = \sigma(\bigcup \mathscr{F}_n)$, 其中 $\mathscr{F}_n = \sigma(x_0, \cdots, x_n), n \geqslant 0$.

注 在"最优停止规则的一般理论中"完全不要求 Ω 是坐标空间. 然而, 就是在"一般理论中"仍然需要假设 Ω 是充分"丰富"的 (详见 [78]).

对于我们来讲, 假设"坐标性"可以简化讨论, 特别对于广义马尔可夫性情形 (§2 定理 1), 正是在此假设下进行的.

像上一节一样, 以 $P(x;B)$ 表示所考虑的链的转移函数: $P(x;B) = \mathbf{P}_x\{X_1 \in B\}, x \in E, B \in \mathscr{E}$.

设 T 是作用于 $\mathscr{E}$– 可测函数 $f = f(x)$ 的一步转移算子,

$$(Tf)(x) = \mathbf{E}_x f(x_1) \left(= \int_E f(y)P(x;\mathrm{d}y) \right), \tag{1}$$

假设其中 $\mathbf{E}_x|f(x_1)| < \infty, x \in E$. (为简便计, 记号 $(Tf)(x)$ 往往写成 $Tf(x)$. 在其他相近的情形下, 也采用类似的约定.)

3. 最优停时 为表述马尔可夫链 X 的最优停时问题, 假设给定某一 $\mathscr{E}$– 可测实函数 $g = g(x)$, 满足条件: $\mathbf{E}_x|g(x_n)| < \infty, x \in E$, 对一切 $n \geqslant 0$ (或 $0 \leqslant n \leqslant N$, 如果事先存在某个"最后的"值 N, 而且在 N 之前应得到"最优解").

以 $\mathfrak{M}_0^n$ 表示 (关于过滤 $(\mathscr{F}_k)_{0\leqslant k\leqslant N}$ 的) 马尔可夫时间 $\tau = \tau(\omega)$ 类, 且在"停时"集合 $\{0,1,\cdots,n\}$ 上取值.

下面的定理是, 第七章 §13 中定理 1 和定理 2 的"马尔可夫"提法.

定理 1 设对于 $0 \leqslant n \leqslant N$ 和 $x \in E$, "(价) 值"为

$$s_n(x) = \sup_{\tau\in\mathfrak{M}_0^n} \mathbf{E}_x g(X_\tau), \tag{2}$$

其中 $\mathbf{E}_x$ 表示对测度 $\mathbf{P}_x$ 求平均.

假设

$$\tau_0^n = \min\{0 \leqslant k \leqslant n : s_{n-k}(X_k) = g(X_k)\}, \tag{3}$$

而

$$Qg(x) = \max\{g(x), Tg(x)\}. \tag{4}$$

那么, 下面的命题成立,

1) 时间 τ_0^n 在类 $\mathfrak{M}_0^n$ 中是最优停时: 对于一切 $x \in E$,

$$\mathbf{E}_x g(X_{\tau_0^n}) = s_n(x). \tag{5}$$

2) 函数 $s_n(x)$ 可以按如下公式来求:

$$s_n(x) = Q^n g(x), \quad x \in E, \tag{6}$$

其中对于 $n = 0, Q^0 g(x) = g(x)$.

3) 函数 $s_n(x) (n \leqslant N)$ 满足递推关系式 $(s_0(x) = g(x))$:

$$s_n(x) = \max\{g(x), Ts_{n-1}(x)\}, \quad x \in E, \quad 1 \leqslant n \leqslant N. \tag{7}$$

证明 将第七章 §13 定理 1 和 2 的结果, 用于函数 $f_n = g(X_n), 0 \leqslant n \leqslant N$.

为此固定某个 "初始" 状态 $x \in E$, 并且考虑上面提到的 §13 中引进的函数 V_n^N 和 v_n^N. 这时, 为强调与初始状态有关, 我们特别使用记号 $V_n^N = V_n^N(x)$. 这样,

$$V_n^N(x) = \sup_{\tau \in \mathfrak{M}_n^N} \mathbf{E}_x g(X_\tau), \tag{8}$$

其中 $\mathfrak{M}_0^n$ 表示 (关于过滤 $(\mathscr{F}_k)_{k \leqslant N}$ 的) 所有马尔可夫时间 $\tau = \tau(\omega)$ 类, 且在 "停时" 集合 $\{n, n+1, \cdots, N\}$ 上取值.

由于第七章 §13 的 (6) 式, 函数 v_n^N 决定于如下递推公式:

$$v_N^N = g(X_N), \quad v_n^N = \max\left[g(X_n), \mathbf{E}_x(v_{n+1}^N | \mathscr{F}_n)\right]. \tag{9}$$

由于广义马尔可夫性 (§2 定理 1), ($\mathbf{P}_x$ – a.c.) 有

$$\mathbf{E}_x(v_N^N | \mathscr{F}_{N-1}) = \mathbf{E}_x(g(X_N) | \mathscr{F}_{N-1}) = \mathbf{E}_{X_{N-1}} g(X_1), \tag{10}$$

其中 $\mathbf{E}_{X_{N-1}} g(X_1)$ 应作如下理解 (见 §2): 取函数 $\psi(x) = \mathbf{E}_x g(X_1)$, 即 $\psi(x) = (Tg)(x)$, 而根据定义认为

$$\mathbf{E}_{X_{N-1}} g(X_1) \equiv \psi(X_{N-1}) = (Tg)(X_{N-1}).$$

这样, $v_N^N = g(X_N)$, 而

$$v_{N-1}^N = \max[g(X_{N-1}), (Tg)(X_{N-1})] = (Qg)(X_{N-1}). \tag{11}$$

类似地继续推导, 可见对于任意 $0 \leqslant n \leqslant N-1$, 有

$$v_n^N = (Q^{N-n} g)(X_n), \tag{12}$$

特别

$$v_0^N = (Q^N g)(X_0) = (Q^N g)(x), \quad (\mathbf{P}_x - \text{a.c.}).$$

根据第七章 §13 的 (13) 式, $v_0^N = V_0^N$. 由于 $V_0^N = V_0^N(x) = s_N(x)$, 可见 $s_N(x) = (Q^N g)(x)$. 于是, 对于 $n = N$ 就证明了 (6) 式. 类似地, 对于 $n < N$ 可以证明 (6) 式.

由 (6) 式和算子 Q 的定义可得 (7) 式.

现在证明, 当 $n = N$ 时, 由 (3) 式定义的停时 τ_0^n 在类 $\mathfrak{M}_0^N$ 中是最优的. 同样, 对于 $n < N, \tau_0^n$ 在类 $\mathfrak{M}_0^n$ 中也是最优的.

根据第七章 §13 的定理 1, 最优停时为

$$\tau_0^N = \min\{0 \leqslant k \leqslant N : v_k^N = g(X_k)\}.$$

由 (12) 式以及上面证明的事实: 对于任意 $n \geqslant 0, s_n(x) = (Q^n g)(x)$, 可见

$$v_k^N = (Q^{N-k} g)(X_k) = s_{N-k}(X_k), \tag{13}$$

从而

$$\tau_0^N = \min\{0 \leqslant k \leqslant N : s_{N-k}(X_k) = g(X_k)\}. \tag{14}$$

于是, 停时 τ_0^n 在类 $\mathfrak{M}_0^N$ 中的最优性得证. □

4. 停止区域和继续观测区域 记

$$\mathbb{D}_k^N = \{x \in E : s_{N-k}(x) = g(x)\}. \tag{15}$$

$$\mathbb{C}_n^k = E \setminus \mathbb{D}_k^N = \{x \in E : s_{N-k}(x) > g(x)\}. \tag{16}$$

那么, 由 (14) 式, 可见

$$\tau_0^N(\omega) = \min\{0 \leqslant k \leqslant N : X_k(\omega) \in \mathbb{D}_k^N\}, \tag{17}$$

而与第七章 §13 的第 6 小节引进的、 Ω 中的集合 D_k^N 和 C_k^N 类似, 集合

$$\mathbb{D}_0^N \subseteq \mathbb{D}_1^N \subseteq \cdots \subseteq \mathbb{D}_N^N = E, \tag{18}$$

$$\mathbb{C}_0^N \supseteq \mathbb{C}_1^N \supseteq \cdots \supseteq \mathbb{C}_N^N = \varnothing, \tag{19}$$

可以相应地称为 E 中的 "停止" 区域和 "继续" 观测区域.

我们指出, 所讨论的马尔可夫链的最优停时问题的特点. 与一般情形不同, 在马尔可夫链的情形下, 对 "停止观测还是继续观测" 问题的回答, 决定于马尔可夫链本身的状态 ($\tau_0^N(\omega) = \min\{0 \leqslant k \leqslant N : X_k(\omega) \in \mathbb{D}_k^N\}$), 换句话说, 决定于游动的 "质点" 所处的位置. 这时, 从原则的观点出发, 最优停时问题的完全解决 (即描绘 "(价) 值" $s_N(x)$ 和最优停时 τ_0^N), 在于从递推 "动态规划方程" (7) 式依次寻找函数 $s_0(x) = g(x), s_1(x), \cdots, s_N(x)$.

5. 类 $\mathfrak{M}_0^\infty$ 中的最优停时 设 $\mathfrak{M}_0^\infty$ 是一切有限马尔可夫停时类. 现在, 在假设 $\tau \in \mathfrak{M}_0^\infty$ 的条件下, 讨论最优停时问题.(对于所有 $\omega \in \Omega$, 当 $\tau \in \mathfrak{M}_0^N$ 时, 停时 $\tau \leqslant N$; 当 $\tau \in \mathfrak{M}_0^\infty$ 时, 停时 $\tau = \tau(\omega) < \infty$.)

这样, 假设 "(价) 值"

$$s(x) = \sup_{\tau \in \mathfrak{M}_0^\infty} \mathbf{E}_x g(X_\tau). \tag{20}$$

为使这里不出现数学期望 $\mathbf{E}_x g(X_\tau)$ 的存在性问题, 例如可以假设

$$\mathbf{E}_x \left[\sup_n g^-(X_n)\right] < \infty, \quad x \in E. \tag{21}$$

显然, 这样以来, 该式一定成立, 如果函数 $g = g(x)$ 有界 ($|g(x)| \leqslant C, x \in E$); 特别, 如果链的状态空间有限, 则条件 (21) 成立.

由 "(价) 值" $s_N(x)$ 和 $s(x)$ 的定义, 可见对于一切 $x \in E$, 有

$$s_N(x) \leqslant s_{N+1}(x) \leqslant \cdots \leqslant s(x). \tag{22}$$

自然, 应想到 $\lim\limits_{N\to\infty} s_N(x)$ 等于 $s(x)$. 而假如是这样, 则在 (7) 式中求极限, 将要得到的 “(价) 值” $s(x)$ 应满足方程:

$$s(x)=\max\{g(x),Ts(x)\},\quad x\in E. \tag{23}$$

顺便指出, 由该方程可见, 对于 $s(x),x\in E$, 有 “变分不等式”:

$$s(x)\geqslant g(x), \tag{24}$$

$$s(x)\geqslant Ts(x). \tag{25}$$

不等式 (24) 说明, “(价) 值” $s(x)$ 是 $g(x)$ 的控制函数. 第二个不等式 (25) 说明, 根据马尔可夫过程一般理论的定义, 函数 $s(x)$ 是**峰态的或上调和的**.

这样, 假如可以证明函数 $s(x)$ 满足 (23) 式, 则可以得出结论, “(价)值” $s(x)$ 是 $g(x)$ **峰态强函数**.

我们现在指出如下情形. 假如某一函数 $v(x)$ 是函数 $g(x)$ 峰态强函数, 那么显然有 “变分不等式”:

$$v(x)\geqslant \max\{g(x),Tv(x)\},\quad x\in E. \tag{26}$$

不过, 结果表明, 如果补充假设, 函数 $v(x)$ 是**最小峰态强函数**, 则在 (26) 式中将是等式, 即 $v(x)$ 满足方程:

$$v(x)=\max\{g(x),Tv(x)\},\quad x\in E. \tag{27}$$

引理 1　*函数 $g(x)$ 的任何最小峰态强函数 $v(x)$, 都满足方程* (27).

证明　证明相当简单. 显然 $v(x)$ 满足不等式 (26). 记 $v_1(x)=\max\{g(x),Tv(x)\}$. 由于 $v_1(x)\geqslant g(x)$ 和 $v_1(x)\leqslant v(x)$, 可见

$$Tv_1(x)\leqslant Tv(x)\leqslant \max\{g(x),Tv(x)\}=v_1(x).$$

从而 $v_1(x)$ 是 $g(x)$ 的峰态强函数. 由于 $v(x)$ 是最小峰态强函数, 故 $v(x)\leqslant v_1(x)$, 即 $v(x)\leqslant \max\{g(x),Tv(x)\}$. 于是, 这连同不等式 (26) 就证明了等式 (27). □

所进行的基于假设 $s(x)=\lim\limits_{N\to\infty} s_N(x)$ 的预先讨论, 并且导出了 (23) 式, 以及引理 1 的命题提示表征 “(价) 值” 的途径, 看来 $s(x)$ 是函数 $g(x)$ 的**最小峰态强函数**.

而事实上确有下面的定理.

定理 2　*设函数 $g=g(x)$ 满足条件* (21):

$$\mathbf{E}_x\left[\sup_n g^-(X_n)\right]<\infty,\quad x\in E.$$

那么, 下列各命题成立:

(a) *价值 $s(x)$ 是函数 $g(x)$ 的最小峰态强函数.*

(b) 价值 $s(x)$ 等于 $\lim\limits_{N\to\infty} s_N(x) = \lim\limits_{N\to\infty} Q^N g(x)$, 并且满足 "瓦尔德 – 贝尔曼 (A. Wald – R. E. Bellman) 动态规划方程"

$$s(x) = \max\{g(x), Ts(x)\}, \quad x \in E.$$

(c) 如果

$$\mathbf{E}_x\left[\sup_n |g(X_n)|\right] < \infty, \quad x \in E,$$

则对于每一个 $\varepsilon > 0$, 停时

$$\tau_\varepsilon^* = \inf\{n \geqslant 0 : s(X_n) \leqslant g(X_n) + \varepsilon\}$$

在类 $\mathfrak{M}_0^\infty$ 中是 ε– 最优的, 即

$$s(x) - \varepsilon \leqslant \mathbf{E}_x g(X_{\tau_\varepsilon^*}), \quad x \in E.$$

如果 $\mathbf{P}_x\{\tau_0^* < \infty\} = 1, x \in E$, 则 τ_0^* 是最优的 (0 – 最优的), 即

$$s(x) = \mathbf{E}_x g(X_{\tau_0^*}), \quad x \in E. \tag{28}$$

(d) 如果集合 E 有限, 则停时 τ_0^* 属于 $\mathfrak{M}_0^\infty$, 并且是最优的.

注 完全可能, 对于某个状态 $x \in E$, 停时 $\tau_0^* = \inf\{n \geqslant 0 : s(X_n) = g(X_n)\}$ 以正概率等于 $+\infty, \mathbf{P}_x\{\tau_0^* = \infty\} > 0$. (甚至有可能对于可数个状态出现这样的情形; 练习题 1.) 因而, 需要约定, 在 "值 X_∞" 无定义的情况下, 应如何理解 "停时 τ取 $+\infty$ 为值".

常规定 $g(X_\infty) \equiv \overline{\lim\limits_n} g(X_n)$ (见第七章 §13 第 1 小节; 亦见 [78]). 还有其他可能的解决办法: 用 $g(X_\tau)I(\tau < \infty)$ 代替 $g(X_\tau)$. 那么, 如果以 $\overline{\mathfrak{M}}_0^\infty$ 表示一切马尔可夫时间类, 其中有可能包含取 $+\infty$ 为值者, 则 "(价) 值"

$$\overline{s}(x) = \sup_{\tau \in \overline{\mathfrak{M}}_0^\infty} \mathbf{E}_x g(X_\tau) I(\tau < \infty) \tag{29}$$

有定义. 从而, 就可以在类 $\overline{\mathfrak{M}}_0^\infty$ 中讨论最优停时问题.

定理 2 的证明 我们仅对于有限集合 E 进行证明. 这种情形比较简单, 而且可以很好地在最优停时问题中显示峰态函数的产生. 关于一般情形的证明, 参见 [78], [102].

(a) 证明 $s(x)$ 是峰态函数, 即 $s(x) \geqslant Ts(x), x \in E$.

显然, 对于每一个状态 $y \in E$ 和 $\varepsilon > 0$, 存在 (一般依赖于 $\varepsilon > 0$ 的) ($\mathbf{P}$ – a. c.) 有限的停时 $\tau_y \in \mathfrak{M}_0^\infty$, 使

$$\mathbf{E}_y g(X_{\tau_y}) \geqslant s(y) - \varepsilon. \tag{30}$$

根据这一停时 $\tau_y, y \in E$, 建立一新停时 $\widehat{\tau}$: 形象地说, 使之决定下一步选择“停时”的“策略”.

假设开始时“质点”位于状态 E. “质点”在此状态不会“逗留”, 然而却实现了一次观测. 设在时刻 $n=1$, “质点”处于状态 $y \in E$. 那么, 由 $\widehat{\tau}$ 表征的“策略”在于, 认为“质点的生命”仿佛重新开始, 而决定其停止的规则, 由停时 τ_y 控制.

而停时 $\widehat{\tau}$ 形式上由如下方式决定.

设 $y \in E$. 考虑事件 $\{\omega : \tau_y(\omega) = n\}, n \geqslant 0$. 由于 τ_y 是马尔可夫时间, 则该事件属于 $\mathscr{F}_n$. 假设 Ω 是由序列 $\omega = (x_0, x_1, \cdots), x_i \in E$, 生成的坐标空间, 而 $\mathscr{F}_n = \sigma(\omega : x_0, \cdots, x_n), x \in E$. 由此可见, 集合 $\{\omega : \tau_y(\omega) = n\}$ 可以写成 $\{\omega : (X_0(\omega), \cdots, X_n(\omega)) \in B_y(n)\}$, 其中 $B_y(n)$ 是 $\mathscr{E}^{n+1} = \mathscr{E} \otimes \cdots \otimes \mathscr{E}$ ($n+1$个 $\mathscr{E}$) 中的某个集合. (亦见第二章 §2 的定理 4.)

停时 $\widehat{\tau} = \widehat{\tau}(\omega)$ 是这样定义的, 其值形如 $n+1 (n \geqslant 0)$, 这时, 在集合

$$\widehat{A}_n = \sum_{y \in E} \{\omega : X_1(\omega) = y, (X_1(\omega), \cdots, X_{n+1}(\omega)) \in B_y(n)\}$$

上, $\widehat{\tau}(\omega) = n+1$. (直观上, 关于停时 $\widehat{\tau}$ 可以作如下说明: 对于任何状态 x, 在时刻 $n=0$肯定进行观测; 假如这时 $X_1 = y$, 则下一步使用停时 τ_y.)

由于 $\sum\limits_{n \geqslant 0} \widehat{A}_n = \Omega$, 则停时 $\widehat{\tau} = \widehat{\tau}(\omega)$ 确实对于一切 $\omega \in \Omega$ 定义, 并且是马尔可夫的 (练习题 2).

由此构造, 广义马尔可夫性 (§2 中 (2) 式) 和 (30) 式可见, 对任意 $x \in E$, 有

$$\begin{aligned}
\mathbf{E}_x g(X_{\widehat{\tau}}) &= \sum_{n \geqslant 0} \sum_{y \in E} \sum_{z \in E} \mathbf{P}_x\{X_1 = y, (X_1, \cdots, X_{n+1}) \in B_y(n), X_{n+1} = z\} g(z) \\
&= \sum_{n \geqslant 0} \sum_{y \in E} \sum_{z \in E} p_{xy} \mathbf{P}_y\{X_0 = y, (X_0, \cdots, X_n) \in B_y(n), X_n = z\} g(z) \\
&= \sum_{n \geqslant 0} \sum_{y \in E} \sum_{z \in E} p_{xy} \mathbf{P}_y\{(X_0, \cdots, X_n) \in B_y(n), X_n = z\} g(z) \\
&= \sum_{y \in E} p_{xy} \mathbf{E}_y g(X_{\tau_y}) \geqslant \sum_{y \in E} p_{xy}[s(y) - \varepsilon] = Ts(y) - \varepsilon.
\end{aligned}$$

这样,

$$s(x) = \mathbf{E}_x g(X_{\widehat{\tau}}) \geqslant Ts(y) - \varepsilon, \quad x \in E,$$

而由于 $\varepsilon > 0$ 的任意性, 有

$$s(x) \geqslant Ts(x), \quad x \in E,$$

因此函数 $s = s(x), x \in E$ 的峰态性得证.

由上面证明的峰态性 (上调和性), 立即可以得到下面的重要结果.

系 1 对于任意 $x \in E$, 过程 (序列)

$$s = (s(X_n))_{n \geqslant 0} \tag{31}$$

(关于 $\mathbf{P}_x$ - 概率) 是上鞅.

将由第七章 §2 的定理 1 用于该上鞅, 可见对于任意停时 $\tau \in \mathfrak{M}_0^\infty$, 有如下不等式:

$$s(x) \geqslant \mathbf{E}_x s(X_\tau), \quad x \in E, \tag{32}$$

而且, 如果对于中的两个马尔可夫停时 σ 和 τ, 有 $\sigma \leqslant \tau(\mathbf{P}_x - \text{a.c.}, x \in E)$, 则

$$\mathbf{E}_x s(X_\sigma) \geqslant \mathbf{E}_x s(X_\tau), \quad x \in E. \tag{33}$$

(注意, 由于空间 E 有限, 故对于所考虑的情形, 上述第七章 §2 定理 1 的所有条件都成立).

由 (32) 式, 得

系 2 假设对于 (20) 式的最优停时问题, 函数 $g = g(x), x \in E$, 是峰态的 (上鞅的). 那么, 停时 $\tau_0^* \equiv 0$ 是最优停时.

(b) 证明 $s(x) = \lim\limits_N s_N(x), x \in E$.

因为 $s_N(x) \leqslant s_{N+1}(x)$, 所以极限 $\lim\limits_N s_N(x)$存在, 记作 $\overline{s}(x)$. 由于 E 有限, 且对于 $s_N(x)(N \geqslant 0)$, 有递推关系式,

$$s_N(x) = \max\{g(x), Ts_{N-1}(x)\},$$

则当 $N \to \infty$时求极限, 得

$$\overline{s}(x) = \max\{g(x), T\overline{s}(x)\}.$$

由此可见, $\overline{s}(x)$ 是函数 $g(x)$ 的峰态强函数. 但是, 由于 $\overline{s}(x)$ 是最小峰态强函数, 可见 $s(x) \leqslant \overline{s}(x)$. 此外, 由于对任意 $N \geqslant 0$, 显然 $s_N(x) \leqslant s(x)$, 可见 $\overline{s}(x) \leqslant s(x)$.

这样 $\overline{s}(x) = s(x)$, 从而证明了欲证的定理 2 的命题 (b).

(c,d) 证明命题 (c) 和 (d). 最后, 证明停时

$$\tau_0^* = \inf\{n \geqslant 0 : s(X_n) = g(X_n)\}, \tag{34}$$

即

$$\tau_0^* = \inf\{n \geqslant 0 : X_n \in \mathbb{D}^*\}, \tag{35}$$

首达 (停时) 集合

$$\mathbb{D}^* = \{x \in E : s(x) = g(x)\} \tag{36}$$

的时间, (在集合 E 有限的情形下) 在类 $\mathfrak{M}_0^\infty$ 中是最优的.

为此, 首先注意到, 由于满足 $g(\widetilde{x})=\max\limits_{x\in E}g(x)$ 的 $\widetilde{x}$ 肯定属于 $\mathbb{D}^*$, 可见 $\mathbb{D}^*$ 不空. 由于在状态 $\widetilde{x}$ 下 $s(\widetilde{x})=g(\widetilde{x})$, 则显然在最优策略应该是: 在状态 $\widetilde{x}$ 下立刻 "停止". 也正是这一点决定了停时 τ_0^*.

为从类 $\mathfrak{M}_0^\infty$ 中最优的角度考虑停时 τ_0^*, 首先需要断定此时间 τ_0^* 属于类 $\mathfrak{M}_0^\infty$, 即

$$\mathbf{P}_x\{\tau_0^*<\infty\}=1, x\in E. \tag{37}$$

由所作关于状态集合 E 有限性假设, 确实可以证明这一点. (对于状态集合 E 可数的情形, 一般已经不是这样; 练习题 1).

为了证明上述事实, 我们指出, 我们感兴趣的事件 $\{\tau_0^*=\infty\}$ 等于事件

$$A=\bigcap_{n\geqslant 0}\{X_n\notin\mathbb{D}^*\}.$$

这样, 需要证明对于一切 $x\in E$, 有 $\mathbf{P}_x(A)=0$.

如果 $\mathbb{D}^*=E$, 则这显然.

设 $\mathbb{D}^*\neq E$. 由于集合 E 有限性, 存在 $\alpha>0$, 使对于一切 $y\in E\backslash\mathbb{D}^*, g(y)\leqslant s(y)-\alpha$.

那么, 对于任意 $\tau\in\mathfrak{M}_0^\infty$, 有

$$\begin{aligned}
\mathbf{E}_x g(X_\tau)&=\sum_{n=0}^{\infty}\sum_{y\in E}\mathbf{P}_x\{\tau=n,X_n=y\}g(y)\\
&=\sum_{n=0}^{\infty}\sum_{y\in\mathbb{D}^*}\mathbf{P}_x\{\tau=n,X_n=y\}g(y)+\sum_{n=0}^{\infty}\sum_{y\in E\backslash\mathbb{D}^*}\mathbf{P}_x\{\tau=n,X_n=y\}g(y)\\
&\leqslant\sum_{n=0}^{\infty}\sum_{y\in\mathbb{D}^*}\mathbf{P}_x\{\tau=n,X_n=y\}s(y)+\sum_{n=0}^{\infty}\sum_{y\in E\backslash\mathbb{D}^*}\mathbf{P}_x\{\tau=n,X_n=y\}[s(y)-\alpha]\\
&\leqslant\mathbf{E}_x s(X_\tau)-\alpha\mathbf{P}_x(A)\leqslant s(x)-\alpha\mathbf{P}_x(A),
\end{aligned}\tag{38}$$

其中最后一个不等式, 由函数 $s(x)$ 的峰态性 (上鞅性), 以及对其成立的不等式 (32) 得到. 在 (38) 式左侧对 $\tau\in\mathfrak{M}_0^\infty$ 取上确界, 得不等式

$$s(x)\leqslant s(x)-\alpha\mathbf{P}_x(A),\quad x\in E.$$

但是 $|s(x)|<\infty,\alpha>0$. 因此 $\mathbf{P}_x(A)=0, x\in E$, 从而就证明了停时 τ_0^* 的有限性.

现在证明停时 τ_0^* 在类 $\mathfrak{M}_0^\infty$ 中的最优性.

根据 τ_0^* 的定义

$$s(X_{\tau_0^*})=g(X_{\tau_0^*}), \tag{39}$$

鉴于这一性质, 我们考虑函数 $\gamma(x)=\mathbf{E}_x g(X_{\tau_0^*})=\mathbf{E}_x s(X_{\tau_0^*})$. 下面证明这一函数 $\gamma(x)$

1) 是峰态的;

2) 控制函数 $g(x): g(x) \leqslant \gamma(x), x \in E$;

3) $\gamma(x) \leqslant s(x)$,

其中最后一个不等式显然.

由 1) 和 2) 可见, $\gamma(x)$控制函数 $s(x)$, 同时又是函数 $g(x)$ 的最小峰态强函数. 因此, 由 3) 可见 $\gamma(x) = s(x), x \in E$, 从而

$$s(x) = \mathbf{E}_x g(X_{\tau_0^*}), \quad x \in E$$

由此可证 τ_0^* 在类 $\mathfrak{M}_0^\infty$ 中的最优性.

现在来证性质 1) 记 $\overline{\tau} = \inf\{n \geqslant 1 : X_n \in \mathbb{D}^*\}$. 该时刻是马尔可夫停时, $\tau_0^* \leqslant \overline{\tau}, \overline{\tau} \in \mathfrak{M}_1^\infty$, 由于函数 $s(x)$ 是峰态的, 故由性质 (33) 得

$$\mathbf{E}_x s(X_{\overline{\tau}}) \leqslant \mathbf{E}_x s(X_{\tau_0^*}), x \in E, \tag{40}$$

此外, 由于广义马尔可夫性 (见 §2 定理 1 的 (2) 式), 可见

$$\begin{aligned}
\mathbf{E}_x s(X_{\overline{\tau}}) &= \sum_{n=1}^{\infty} \sum_{y \in \mathbb{D}^*} \mathbf{P}_x\{X_1 \notin \mathbb{D}^*, \cdots, X_{n-1} \notin \mathbb{D}^*, X_n = y\} s(y) \\
&= \sum_{n=1}^{\infty} \sum_{y \in \mathbb{D}^*} \sum_{z \in E} p_{xz} \mathbf{P}_z\{X_0 \notin \mathbb{D}^*, \cdots, X_{n-2} \notin \mathbb{D}^*, X_{n-1} = y\} s(y) \\
&= \sum_{z \in E} p_{xz} \mathbf{E}_z s(X_{\tau_0^*}).
\end{aligned} \tag{41}$$

因此, 由 (40) 式, 可见

$$\mathbf{E}_x s(X_{\tau_0^*}) \geqslant \sum_{z \in E} p_{xz} \mathbf{E}_z s(X_{\tau_0^*}),$$

即

$$\gamma(x) \geqslant \sum_{z \in E} p_{xz} \gamma(z), \quad x \in E,$$

这就证明了函数的 $\gamma(x)$ 峰态性.

只剩下证明函数 $\gamma(x)$ 是 $g(x)$ 的强函数.

如果 $x \in \mathbb{D}^*$, 则 $\tau_0^* = 0$, 且显然 $\gamma(x) = \mathbf{E}_x g(X_{\tau_0^*}) = g(x)$.

考虑集合 $E \backslash \mathbb{D}^*$, 并且设 $E_0^* = \{x \in E \backslash \mathbb{D}^* : \gamma(x) < g(x)\}$. 这一集合 E_0^* 是有限的. 设 x_0^* 是函数 $g(x) - \gamma(x)$ 在集合 E_0^*上的极大值点:

$$g(x_0^*) - \gamma(x_0^*) = \max_{x \in E_0^*} [g(x) - \gamma(x)].$$

引进一新函数

$$\widetilde{\gamma}(x) = \gamma(x) + [g(x_0^*) - \gamma(x_0^*)], \quad x \in E. \tag{42}$$

该函数 (作为峰态函数与常数之和) 显然是峰态的, 并且对于一切 $x \in E$,

$$\widetilde{\gamma}(x) - g(x) = [g(x_0^*) - \gamma(x_0^*)] - [g(x) - \gamma(x)] \geqslant 0.$$

这样, 函数 $\widetilde{\gamma}(x)$ 是 $g(x)$ 的峰态强函数. 由于函数 $s(x)$ 是 $g(x)$ 的最小峰态强函数, 故 $\widetilde{\gamma}(x) \geqslant s(x)$.

由此可见,

$$\widetilde{\gamma}(x_0^*) \geqslant s(x_0^*).$$

但是, 由于根据 (42) 式 $\widetilde{\gamma}(x_0^*) = g(x_0^*)$, 故 $g(x_0^*) \geqslant s(x_0^*)$. 因为对于一切 $x \in E, s(x) \geqslant g(x)$, 所以 $g(x_0^*) = s(x_0^*)$, 说明点 x_0^* 属于集合 $\mathbb{D}^*$. 然而, 根据假设 $x_0^* \in E \backslash \mathbb{D}^*$.

所得矛盾说明集合 $E \backslash \mathbb{D}^* = \varnothing$. 由此可见, 对于一切 $x \in E, \gamma(x) \geqslant g(x)$. □

6. 例 现在举几个例子.

例 1 考虑 §8 例 5 中描绘的、具有两个 "吸收" 状态 0 和 N 的简单随机游动. 这时, 我们假设 $p = q = 1/2$ (对称游动). 对于所考虑的随机游动, 如果函数 $\gamma(x), x \in E = \{0, 1, \cdots, N\}$, 是峰态的, 则对于一切 $x = 1, \cdots, N-1$, 有

$$\gamma(x) \geqslant \frac{1}{2}\gamma(x-1) + \frac{1}{2}\gamma(x+1). \tag{43}$$

设给定一函数 $g = g(x), x \in \{0, 1, \cdots, N\}$. 由于 0 和 N 是吸收状态可见, 应当在满足条件 (43) 和边界条件 $\gamma(0) = g(0), \gamma(N) = g(N)$ 的所有函数 $\gamma(x)$ 中求函数 $s(x)$.

条件 (43) 表示函数 $\gamma(x)$ (在集合 $\{1, 2, \cdots, N-1\}$ 上) 的凸性. 因此, 可以得出如下结论: 在该问题中, "(价) 值" $s(x)$, 其中

$$s(x) = \sup_{\tau \in \mathfrak{M}_0^\infty} \mathbf{E}_x g(X_\tau)$$

是最小凸函数, 满足边界条件 $\gamma(0) = g(0), \gamma(N) = g(N)$. 直观上, 为求函数 $s(x)$的值, 需要按如下方式进行. 在函数 $g(x)$ 的值上 "栓" 上 "绷紧的线". 在图 42 "绷紧的线" 通过 $(0, a), (1, b), (4, c), (6, d)$ 4 个点, 其中 0,1,4,6 构成的 $\mathbb{D}^*$ 剩余集, 且在这些点上 $s(x) = g(x)$. 在其余状态 $x = 2, 3, 5$ 上, 要求 $s(x)$ 的值决定于线性内插. "凸流形" $s(x)$ 的值, 对于属于 E 的一切状态 x, 在一般情形下类似地决定.

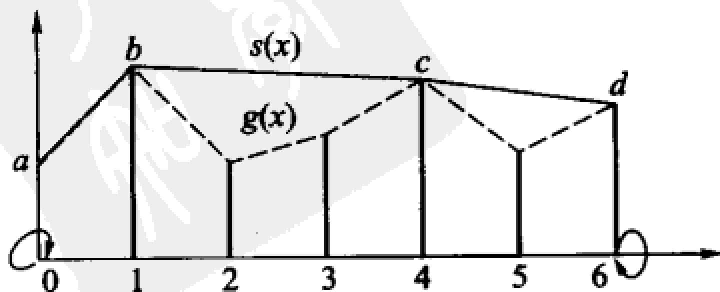


图 42 函数 $g(x)$ (虚线) 及其凸流形 $s(x), x = 0, 1, \cdots, 6$.

例 2 像 §8 中例 7 和例 8 一样, 考虑在状态集合 $E=\{0,1,\cdots,N\}$ 上且在 0 和 N 是反射屏幕的、对称 $(p=q=1/2)$ 简单随机游动. 所考虑的游动是正常返的. 由此可见, 在此最优停时问题中,

$$s(x)=\sup_{\tau\in\mathfrak{M}_0^\infty}\mathbf{E}_x g(X_\tau)$$

具有相当简单和自然的构造: 需要等到 "函数 $g(x)$ 达到最大值的任何一个状态" 的时刻, 并且在此时刻停止观测.

例 3 假设对于在集合 $E=\{0,1,\cdots,N\}$ 上的、简单对称随机游动, 0 是吸收状态而 N 是反射状态. 设 x_0是函数 $g(x)$ 达到最大值, 而且是游动离 N 最近的状态. 那么, 最优停时具有如下形式: 如果 $x_0\leqslant x\leqslant N$, 则当 (以 $\mathbf{P}_x$– 概率 1) 达到状态 x_0 时游动停止.

如果假设 $E=\{0,1,\cdots,x_0\}$, 而 x_0 是吸收状态, 则在状态 0 和 x_0 之间停止问题的解与例 1 一样.

7. 最优对象的选择问题 最后讨论有广泛知名度的 "最优对象的选择问题", 亦称 "选择审慎的未婚女问题", "选秘书问题", …… (参阅 [59], [78], [102], [106]). 为叙述的直观计, 我们选择 "选择审慎的未婚女问题" 的形式.

假设 "未婚女" 希望自 N 个候选人中选出最好的 "未婚夫". 假设 N 事先已知, 并且所有候选人事先已经按 "素质" 排序. 为确定计, 假设序号最大者, 即序号为 N 者是最优秀的, 按 "素质" 排第二位者的序号为 $N-1,\cdots$, 最差者序号为 1.

候选 "未婚夫" 按*随机*的顺序会见 "未婚女", 并且按如下方式实现.

设 $(a_1,a_2,\cdots,a_N)$ 是 $1,2,\cdots,N$ 的排列. 排列的总数等于 $N!$, 并且假设它们都是 "随机" 的, 即每一个排列出现的概率都等于 $1/N!$.

在该问题中, 样本 $(a_1,a_2,\cdots,a_N)$ 的*有序性*在于, (在会见全部候选人的潜在可能性的情况下) 第一个会见 "未婚女" 者是编号为 a_1 的候选人, 然后是 $a_2,\cdots$, 最后是编号为 a_N 的候选人.

加在 "未婚女" 可能 "策略" 由如下的想法形成.

"未婚女" 并 "不了解" 她会见的 "未婚夫" 候选人的完全的素质. 她关于候选人的全部了解, 仅仅是通过*两两比较*的结果, 知道哪个较好或较坏.

其次, 如果 "未婚女" 拒绝了某个 "未婚夫" 候选人, 则他就*不再*与她会面 (然而被拒绝者实际上可能是最优秀的).

"未婚女" 的策略应该是, 根据顺序会见候选人的结果 (并记忆; 两两比较的结果并考虑 "素质"), 这样来选择停时 τ^*, 使得

$$\mathbf{P}\{a_{\tau^*}=N\}=\sup_\tau\mathbf{P}\{a_\tau=N\},\tag{44}$$

其中 τ 属于某一停时类 $\mathfrak{M}_1^\infty$, 而类 $\mathfrak{M}_1^\infty$ 决定于由 "未婚女" 在会见 "未婚夫" 候选人的过程中得到的 "信息".

为更确切地描绘所考虑的停时类 $\mathfrak{M}_1^\infty$, 我们根据序列 $\omega=(a_1,a_2,\cdots,a_N)$ 建立 "秩" 序列 $X=(X_1,X_2,\cdots)$, 该序列是由上面描绘的 "未婚女" 的行动方式自然产生的.

具体地说, 设 $X_1=1$, 并设 X_2 是 "未婚夫" 的大于以上所有编号的编号 (即到达的时刻). 这样, 如果 $X_2=3$, 则说明对于所讨论的序列 $\omega=(a_1,a_2,\cdots,a_N)$, 值 $a_1>a_2$, 而 $a_3>a_1(>a_2)$. 类似地继续定义 $X_3,X_4,\cdots$, 例如设 $X_3=5$. 因而 $a_3>a_4$, 而 $a_5>a_3(>a_4)$.

最多有 N 个强函数 (在 $(a_1,a_2,\cdots,a_N)=(1,2,\cdots,N)$ 的情形下). 假如对于序列 $\omega=(a_1,a_2,\cdots,a_N)$, 强函数的个数等于 m, 则设 $X_{m+1}=X_{m+2}=\cdots=N+1$.

所考虑的停时类 $\mathfrak{M}_1^\infty$, 由具有性质

$$\{\omega:\tau(\omega)=n\}\in\mathscr{F}_n^X$$

的停时 $\tau=\tau(\omega)$ 构成, 其中 $\mathscr{F}_n^X=\sigma(X_1,\cdots,X_n),1\leqslant n\leqslant N$.

现在比较详细地讨论 "秩" 序列 $X=(X_1,X_2,\cdots)$ 的构造.

不难证明 (练习题 3), 该序列是相空间为 $E=\{1,2,\cdots,N+1\}$ 齐次马尔可夫链. 该链的转移概率决定于公式:

$$p_{ij}=\frac{i}{j(j-1)},\quad 1\leqslant i<j\leqslant N,\tag{45}$$

$$p_{i,N+1}=\frac{i}{N},\quad 1\leqslant i\leqslant N,\tag{46}$$

$$p_{N+1,N+1}=1.\tag{47}$$

由此可见, $N+1$ 是吸收状态, 而且只有沿集合 E "向上" 转移是可能的, 即只可能转移 $i\to j$, 并且 $j>i$.

注 考虑到每一个序列 $\omega=(a_1,a_2,\cdots,a_N)$ 出现的概率都等于 $1/N!$, 故可由如下简单地讨论得到 (45) 式.

对于 $1\leqslant i<j\leqslant N$, 转移概率为

$$p_{ij}=\mathbf{P}\{X_{n+1}=j|X_n=i\}=\frac{\mathbf{P}\{X_n=i,X_{n+1}=j\}}{\mathbf{P}\{X_n=i\}}.\tag{48}$$

事件 $\{X_n=i,X_{n+1}=j\}$ 表示, a_j的值在 $a_1,\cdots,a_j$ 各值中最大的, 这时 $a_j>a_i$. 上述事件的概率等于

$$\mathbf{P}\{X_n=i,X_{n+1}=j\}=\frac{(j-2)!}{j!}=\frac{1}{j(j-1)}.$$

同样, 事件 $\{X_n=i\}$ 表示, a_i的值在 $a_1,\cdots,a_i$各值中是最大的, 而事件 $\{X_n=i\}$ 的概率等于

$$\mathbf{P}\{X_n=i\}=\frac{(i-1)!}{i!}=\frac{1}{i}.$$

由这些讨论和 (48) 式, 得 (45) 式.

为证明 (46) 式, 只需注意到, 如果 $X_n = i$, 则 $X_{n+1} = N+1$, 说明 a_i的值在 $a_1, \cdots, a_{i-1}$ 各值中, 以及在 $a_{i+1}, \cdots, a_{N+1}$ 中都是最大的.

最后, 公式 (47) 显然.

现在假设 "未婚女" 选择 (关于 σ- 代数系 $(\mathscr{F}_n^X)$ 的) 某一停时 τ, 并且使 $X_\tau = i$. 那么, 根据 (46) 式, 此时间 τ 结果是成功的 (即 $a_\tau = N$) 条件概率等于 $X_\tau / N (= i/N)$. 从而,

$$\mathbf{P}\{a_\tau = N\} = \mathbf{E}\frac{X_\tau}{N}.$$

因此, 求最优停时 τ^*, 即求满足

$$\mathbf{P}\{a_{\tau^*} = N\} = \sup_\tau \mathbf{P}\{a_\tau = N\}$$

的时间 τ^*, 归结为解最优停时问题

$$V^* = \sup_\tau \mathbf{E}\frac{X_\tau}{N}, \tag{49}$$

其中 τ 关于 σ - 代数系 $(\mathscr{F}_n^X)$ 是马尔可夫时间.

在 (49) 式中设 $X_1 = 1$. 根据对马尔可夫序列, 求解最优停时问题的一般方法, 记

$$v(i) = \sup_\tau \mathbf{E}_i g(X_\tau),$$

其中 $\mathbf{E}_i$ 是在 $X_n = i$ 的条件下的数学期望, 且

$$g(i) = \frac{i}{N}, \quad i \leqslant N, \quad g(N+1) = 0.$$

由定理 2 已知, 函数 $v(i), 1 \leqslant i \leqslant N+1$, 是函数 $g(i), 1 \leqslant i \leqslant N+1$, 的**峰态强函数**:

$$v(i) \geqslant Tv(i) = \sum_{j=i+1}^{N} \frac{i}{j(j-1)} v(j), \tag{50}$$

$$v(i) \geqslant g(i), \tag{51}$$

并且 (函数 $v(i)$) 是同类函数中**最小的**. 由同一定理 2, 可见满足方程:

$$v(i) = \max\{g(i), Tv(i)\}, \quad 1 \leqslant i \leqslant N+1. \tag{52}$$

这时, 不难看出, 所求函数 $v(i)$ 应当满足:

$$v(N+1) = 0, \quad v(N) = g(N) = 1.$$

以 $\mathbb{D}^*$ 表示停止观测的状态 $i \in E$ 的集合. 根据定理 1, 该集合可以用如下方式描绘:

$$\mathbb{D}^* = \{i \in E : v(i) = g(i)\}.$$

相应地, 继续观测的集合为

$$\mathbb{C}^* = \{i \in E : v(i) > g(i)\}.$$

这样, 如果 $i \in \mathbb{D}^*$, 则

$$g(i) = v(i) \geqslant Tv(i) = \sum_{j=i+1}^{N} \frac{i}{j(j-1)} v(j) \geqslant \sum_{j=i+1}^{N} \frac{i}{j(j-1)} g(j)$$
$$= \sum_{j=i+1}^{N} \frac{i}{j(j-1)} \cdot \frac{j}{N} = g(i) \sum_{j=i+1}^{N} \frac{1}{j-1}.$$

从而, 如果 $i \in \mathbb{D}^*$, 则应该成立不等式:

$$\sum_{j=i+1}^{N} \frac{1}{j-1} \leqslant 1.$$

此外, 如果该不等式成立, 且值 $i+1, \cdots, N$ 都属于 $\mathbb{D}^*$, 则

$$Tv(i) = \sum_{j=i+1}^{N} \frac{i}{j(j-1)} g(j) = g(i) \sum_{j=i+1}^{N} \frac{1}{j-1} \leqslant g(i),$$

因而, 状态 i 也属于集合 $\mathbb{D}^*$.

由 $v(N) = g(N)$, 可见 $N \in \mathbb{D}^*$. 因此, 上面的讨论说明, 集合 $\mathbb{D}^*$ 应该有如下形式:

$$\mathbb{D}^* = \{i^*, i^*+1, \cdots, N, N+1\},$$

其中 $i^* = i^*(N)$ 决定于不等式

$$\frac{1}{i^*} + \frac{1}{i^*+1} + \cdots + \frac{1}{N-1} \leqslant 1 < \frac{1}{i^*-1} + \frac{1}{i^*} + \cdots + \frac{1}{N-1}, \tag{53}$$

而对于充分大的 N, 由此得

$$i^*(N) \sim \frac{N}{\mathrm{e}}. \tag{54}$$

事实上, 对于任意 $n \geqslant 2$, 有

$$\ln(n+1) - \ln n < \frac{1}{n} < \ln n - \ln(n-1).$$

因此,

$$\ln\frac{N}{n} < \frac{1}{n} + \cdots + \frac{1}{N-1} < \ln\frac{N-1}{n-1},$$

由此连同 (53) 式得不等式

$$\ln\frac{N}{i^*(N)} < 1 < \ln\frac{N-1}{i^*(N)-2}.$$

于是, 由此不等式得渐近式 (54).

现在对于 $i \in E = \{1, 2, \cdots, N+1\}$, 求函数 $v = v(i)$.

如果 $i \in \mathbb{D}^* = \{i^*, i^*+1, \cdots, N, N+1\}$, 则 $v(i) = g(i) = i/N$.

设 $i = i^* - 1$, 则

$$v(i^*-1) = Tv(i^*-1) = \sum_{j=i^*}^{N} \frac{i^*-1}{j(j-1)} g(j) = \frac{i^*-1}{N}\left(\frac{1}{i^*-1} + \cdots + \frac{1}{N-1}\right).$$

设 $i = i^* - 2$, 则

$$\begin{aligned} v(i^*-2) = Tv(i^*-2) &= \frac{i^*-2}{(i^*-1)(i^*-2)} v(i^*-1) + \sum_{j=i^*}^{N} \frac{i^*-2}{j(j-1)} g(j) \\ &= \frac{1}{N}\left(\frac{1}{i^*-1} + \cdots + \frac{1}{N-1}\right) + \frac{i^*-2}{N} \sum_{j=i^*}^{N} \frac{1}{j-1} \\ &= \frac{i^*-1}{N}\left(\frac{1}{i^*-1} + \cdots + \frac{1}{N-1}\right). \end{aligned}$$

由归纳法可得, 对于一切 $1 \leqslant i < i^*$, 有

$$v(i) = v^*(N) = \frac{i^*-1}{N}\left(\frac{1}{i^*-1} + \cdots + \frac{1}{N-1}\right). \tag{55}$$

从而, 对于 $i \in \{1, 2, \cdots, N\}$, 有

$$v(i) = \begin{cases} v^*(N), & 1 \leqslant i < i^*(N), \\ g(i) = \dfrac{i}{N}, & i \leqslant N. \end{cases} \tag{56}$$

注意到 (55) 式, 由于

$$\lim_{N\to\infty}\left(\frac{1}{i^*(N)-1} + \cdots + \frac{1}{N-1}\right) = 1, \tag{57}$$

则当 $N \to \infty$ 时

$$\lim_{N\to\infty} v^*(N) = \lim_{N\to\infty} \frac{i^*(N)-1}{N} = \frac{1}{\mathrm{e}} \approx 0.0368. \tag{58}$$

初看来, 所得结果可能有些奇怪, 因为由此可见, 如果候选人数非常大, 则 “未婚女” 以很大的概率

$$V^* = \sup_{\tau} \mathbf{P}\{a_\tau = N\} = v^*(N) \approx 0.368,$$

存在从他们之中选择最优秀者的策略. 这时, 最优停时为

$$\tau^* = \inf\{n : X_n \in \mathbb{D}^*\},$$

其中 $\mathbb{D}^* = \{i^*, i^*+1, \cdots, N, N+1\}$.

这样, "未婚女" 最优策略是, 观察 i^*-1个候选人, 其中

$$i^* = i^*(N) \sim \frac{N}{\mathrm{e}}, \quad n \to \infty,$$

然后, 在随后会见的候选人中, 选择第一个比以前会见的都优秀者.

在 $N=10$ 的情形下, 较细致地分析表明 (例如, 见 [102] 的第三章 §1), $i^*(10)=4$. 换句话说, 这时需要会见 3 个候选人, 并在随后会见的候选人中, 选择选择第一个比前 3 个都优秀者. 选中最优秀 "未婚夫" 相应的概率 (即 v^* (10) 的值) 等于 0.399.

8. 练习题

1. 举例说明, 对于具有可数状态集的马尔可夫链, (在类 $\mathfrak{M}_0^\infty$ 中) 有可能不存在最优停时.

2. 验证在定理 2 的证明中引进的时间 τ_y, 是马尔可夫时间.

3. 证明在第 7 小节, 讨论 "选择审慎的未婚女问题" 时引进的序列 $X=(X_1, X_2, \cdots)$, 是齐次马尔可夫链.

4. 设 $X=(X_n)_{n\geqslant 0}$ 是在 $\mathbb{R}$ 中取值的齐次马尔可夫链, 而其转移函数为 $P=P(x;B), x\in\mathbb{R}, B\in\mathscr{B}(\mathbb{R})$. 称 $\overline{\mathbb{R}}$- 函数 $f=f(x), x\in\mathbb{R}$, 是 P - 调和函数 (或关于 P 是调和函数), 如果

$$\mathbf{E}_x|f(X_1)| = \int_{\mathbb{R}} |f(y)|P(x;\mathrm{d}y) < \infty, \quad x\in\mathbb{R},$$

而

$$f(x) = \int_{\mathbb{R}} f(y)P(x;\mathrm{d}y), \quad x\in\mathbb{R}. \tag{59}$$

(如果将 (59) 式中的 "=" 换成 "⩾", 则函数 $f=f(x)$ 为上调和函数).

证明如果 f 是上调和函数, 则对于任意 $x\in\mathbb{R}$, 序列 $(f(X_n))_{n\geqslant 0}(X_0=x)$ (关于测度 $\mathbf{P}_x$) 是上鞅.

5. 证明 (38) 式中的停时 τ 属于类 $\mathfrak{M}_1^\infty$.

6. 仿照第 6 小节中的例 1, 对于 §8 的例中的所有简单随机游动, 考虑最优停时问题:

$$s_N(x) = \sup_{\tau\in\mathfrak{M}_0^N} \mathbf{E}_x g(X_\tau)$$

和

$$s(x) = \sup_{\tau\in\mathfrak{M}_0^\infty} \mathbf{E}_x g(X_\tau).$$

概率的数学理论形成的简史

在叙述概率论的历史问题时, 可以程式化地分为如下的几个阶段 (对照 [26] *), [43]):

史前

第一时期 (17 世纪 —18 世纪初)

第二时期 (18 世纪 —19 世纪初)

第三时期 (19 世纪后半叶)

第四时期 (20 世纪初和 20 世纪中叶)

史前. 关于随机性的直观印象,以及关于可能机会的各种不同的论断 (涉及宗教祭祀活动, 解决纠纷, 预测等), 进入世纪的纵深. 在前科学时代, 还涉及一些人类尚未被认识的现象, 以及尚无合理解释的现象, 并且只有在几个世纪之前才开始思考和真正进行逻辑研究的现象.

考古资料表明, 第一个 "随机工具" 的难得器件是, 很久就被用于原始 (极简单) 的博弈的色子 (astragalus) **). 可以肯定地说, 在如下一些年代, 这样的色子在桌上的博弈中已经使用: 埃及的第一个王朝 (大约公元前 3500 年), 之后在古希腊和古罗马. 众所周知 [20], 罗马皇帝奥古斯特 (August, 公元前 63 年 — 公元 14 年), 克劳迪厄斯 (Claudius, 公元前 10 年 — 公元 54 年), 曾经是古怪的色子博弈者.

与博弈相联系, 那时已经出现了关于 "有利结局" 和 "不利结局" 的个数问题. 除博弈之外, 在保险业和商业中出现了类似的问题. 已知保险业最早的形式, 是在巴比伦的记事中发现的海运合同, 大致在公元前 4000 年 — 公元前 3000 年. 后来, 类

*) 这一部分的引文见第 894 — 898 页上文献索引.

**) astragalus 是一种刻有点 "•" 的 "双蹄目" 的骨头, 其形状为: 外形如正四棱柱, 有 4 个平面, 上下底面为球面. 掷到桌面上时, 只有其中一个平面朝上.

似契约的实际经腓尼基① 人传给希腊人, 罗马人, 印度人. 其踪迹可以在罗马文化的早期法典和拜占庭帝国② 的法律中找到. 鉴于寿命保险的需要, 罗马法学家尤尔皮安 (Ulpian) 于 (公元前 220 年) 编制了第一个死亡表. 在意大利的城市 - 共和国 (罗马, 威尼斯, 热那亚, 比萨, 佛罗伦萨), 鉴于保险的实际, 出现了简单统计及实际核算的必要性. 众所周知, 第一份确切注明日期的寿命保险的合同, 于 1347 年在热那亚签订.

城市 - 共和国开创了文艺复兴时期 (Renaissance, 14 世纪末 — 17 世纪初) —— 在西欧的改造和更新时期. 看来, 正是在意大利的文艺复兴时期, 出现了或多或少严重的争议, 主要是 L. 帕乔里 (Luca Pacioli, 1445 年 — 1517 年 (?)), C. 卡利卡尼尼 (Celio Calcagnini, 1479 年 — 1541 年) 和 N. F. 塔尔塔利亚 (Nicola Fontana Tartaglea, 1500 年 — 1557 年) 关于 "概率" 论断的争议, 基本上是哲学性质的争议 (见 [43], [20]).

看来, G. 卡尔达诺 (Gerolamo Cardano, 1501 年 — 1576 年) 是最早开始从数学上分析博弈结局者之一, 他是广为人知的 "卡尔达诺轴" 的发明者, 并且求解了 3 次方程. 他的文献抄本 (大约 1525 年), 只有在 1563 年, 以 "Liber de Ludo Aleæ" (《关于博弈的书》) 书名才出版, 该书并不仅仅是某种赌徒的实际参考书. 在该书中最早阐述了组合的思想, 利用组合方法便于描绘一切可能结局的集合 (在不同次数地, 投掷不同特点的色子时, 可能结局的集合). G. 卡尔达诺还发现, 对于均匀对称的色子, "'有利组合数' 与'一切可能组合总数' 的比值, 与博弈的实际十分一致" [20].

1. 第一时期 (17 世纪 — 18 世纪初)　许多人, 例如, 拉普拉斯 [39] (亦见 [61]) 把《概率演算》(《概率论》) 1654年登记与帕斯卡 (Blaise Pascal,1623 年 — 1662 年) 及费马 (Pierre de Fermat, 1601 年 — 1665 年) 的通信相联系. 通信与勋章获得者 de 梅尔 (Chevalier de Méré, 他也是 Antoine Gombaud —— 作家和道学先生, 1607 年 — 1684 年) 给帕斯卡提出的某些问题有关.

这些问题之一就是, 在中断的博弈中, 如何公平地分配赌注. 具体地说, 假设两个选手 A 和 B 约定, 在对弈中 (例如, 共对弈 5 局) 获胜者将得到全部赌注. 假设当选手 A 获 4 胜而选手 B 获 3 胜时, 则对弈强行停止. 问在此被停止的对弈中, 选手应按什么比例分配赌注? 仿佛该问题的 "自然" 答案之一就是, 应当按 $2:1$ 的比例分配赌注. 实际上, 对弈经两步肯定结束, 这时选手 A 只需赢一局, 而选手 B 需要赢两局. 由此可见, 仍导致 $2:1$ 的比例.

然而, 按局数赢得对弈的选手, "自然" 也可以认为是 $4:3$. 如帕斯卡和费马认为两个都不是: 应当按 $3:1$ 的比例分配赌注.

另一个问题是, 哪一种情形更有可能: 将一枚均匀对称的色子掷 4 次, "6 个点至少出现一次", 还是将两枚均匀对称的色子同时掷 24 次, "数偶 (6,6) 至少出现一次".

①腓尼基 (Phocnicia) 地中海东岸的古国, 曾在地中海开辟了许多殖民地, 公元前6 世纪被波斯征服, 公元前 332 年被马其顿王朝亚历山大占领. —— 译者

②拜占庭 (Byzantium) 帝国, 亦称东罗马帝国 (公元 4 — 5 世纪) 首都君士坦丁堡. —— 译者

对于上述问题, 帕斯卡和费马给出了正确的答案: 第一种情形比第二种情形可能性更大. 两种情形的概率分别等于

$$p_1 = 1 - \left(\frac{5}{6}\right)^4 \approx 0.516 \quad 和 \quad p_2 = 1 - \left(\frac{35}{36}\right)^{24} \approx 0.491.$$

在求解这些问题时, 帕斯卡和费马 (像卡尔达诺一样) 广泛运用了组合分析的方法. 组合分析法, 在计算不同结局的个数时, 成为 “概率演算” 的基本方法之一. 较早就著名的帕斯卡三角形①, 在这里也找到了其 “应用” 的位置.

在 1657 年出版了 C. 惠更斯 (Christianus Huygens, 1629 年 — 1695 年) 的书《De Retiociniis in Lugo Aleæ》(《博弈中的计算》), 被认为是 “概率演算” 的第一部系统的著作. 书中以明显的形式, 表述了许多基本概念, 概率的演算原则, 引进了概率的加法法则和乘法法则, 还包括关于数学期望概念的讨论. 在很常时间里, 该书是《初等概率论》的主要参考书.

所涉及的 “概率论” 形成时期的核心代表人物是, J. 伯努利 (Jacob〈Jakob, James, Jacques〉Benoulli, 1654 年 — 1705 年). J. 伯努利的功勋在于, 他把 “有关所考虑事件的可能结局数” 与 “可能结局的总数” 的比值, 作为 “事件概率” 的 “古典型” 概念引进了科学.

J. 伯努利的与其名字联系在一起的基本成果, 当然是作为概率论一切应用基础的*大数定律*.

J. 伯努利于 1713 年, 将这一定律用极限定理的正式形式, (在其侄子 N. 伯努利 [Nikolaus Benoulli] 的参与下) 出版了的专著《Ars Conjectandi》(《假设的艺术》), 人们把该专著出版的日期当作大数定律的生日 (见 [3] 第 9, 27, 75, 83 页). 像 A. A. 马尔可夫在纪念大数定律 200 周年时的发言中 (见 [53],[3]) 指出的, J. 伯努利在 (1703 年 10 月 3 日和 1704 年 4 月 20 日) 给 G.W. 莱布尼茨 (Gottfried Wilhelm Leibniz, 1646 年 — 1716 年) 信中写到, 他 “在 20 年前已经知道” 这个定理. (“大数定律” 这一术语是泊松 1835 年建议启用的.)

伯努利家族的另一名代表, 是 D. 伯努利 (Daniel Benoulli, 1667 年 — 1748 年), 因为他在概率论中关于所谓 “彼得堡奇论” 的争议而知名. 对于该问题的解, 他使用了 “道德期望” 的概念.

概率论形成的第一个时期, 正是数学科学建立的时代. 使用诸如连续性, 无限大和无限小等概念, 正是属于这个时期. I. 牛顿 (Isaac Newton, 1642 年 — 1727 年) 和 G. W. 莱布尼茨创建微分学与积分学, 也是在这个时期. 像 A. H. 柯尔莫戈洛夫 [26] 指出的那样, 这个时期的任务是, “理解研究因果关系的、数学方法不寻常的广度和深度 (而当时似乎是万能的). 作为根据系统现在的状态, 唯一决定系统将来的变化的定律的微分方程的思想研究, 与现在相比在数学科学中还占有相当特殊的地位. 在

①即杨辉三角形. 杨辉 (中国数学家, 13 世纪, 南宋), 帕斯卡 (17 世纪, 法国数学家). —— 译者

数学科学中, 在微分方程的确定性概型不能应用的地方, 需要概率论. 在当时具体的自然科学的资料, 对于计算和事物性的应用, 尚未用到概率论.

然而, 对于微分方程组一类的确定性模型已经十分明显, 引进实际现象的、深入的模式化是不可避免的. 同样明显的是, 对于无法单独考虑, 且相互之间没有关联的大量现象混乱的基础上, ‘平均’ 来讲可能出现完全清晰的规律性", 这充分地揭示了 J. 伯努利的极限定理 —— 大数定律的意义.

必须指出, J. 伯努利意识到, 考虑重复试验结果的无限序列的重要性. 关于事件在试验中出现频率的*极限性质*提法本身, 在局限于初等算术和简单组合方法的、概率论的研究中曾经是新的 (“无限的”) 思想. 恰好是导致大数定律的、问题这样的提法, 不但显示出, “事件出现的*概率*” 与 “事件在有限次重复试验中出现的*频率*” 之间的差异, 而且显示出, 用试验次数充分大时频率的值, 以一定的准确性确定事件概率的可能性.

2. 第二时期 (18 世纪 — 19 世纪初) 这个时期主要有这样一些代表人物: P. R. 蒙特莫尔特 (Pierre–Remond de Montmort, 1678 年 — 1719 年), Dé. 棣莫弗 (Abraham Dé Moivre, 1667 年 — 1754 年), T. 贝叶斯 (Thomas Bayes, 1702 年 — 1761 年); P. S. 拉普拉斯 (Pierre Simon de Laplace, 1749 年 — 1827 年), C. F. 高斯 (Carl Friedrich Gauss, 1777 年 — 1855 年); S. D. 泊松 (Siméon Denis Poisson, 1781 年 — 1840 年).

如果说第一时期实质上带有*哲学*的特点, 那么在第二时期是发展和*分析*方法的精练. 在各种不同领域出现了进行计算的必要性, 将概率统计方法用于观测误差理论, 射击理论等.

蒙特莫尔特, 以及棣莫弗受到 J. 伯努利在 “概率的演算” 工作的强烈影响. 蒙特莫尔特在他的书《Essai d’ Analyse sur les jeux de Hasard》(《随机博弈试验分析》, 1708 年) 中, 重点就是不同博弈中核算方法的展开.

棣莫弗在他的两部书中:《Doctrine of Chances》(《偶然性的理论》; 1718 年),《Miscellanea Analytica Supplementum》(《分析方法》或《分析混合体》, 1730 年), 相当详细地给出了如下一些定义: 事件的*独立性*, *期望*, *条件概率*.

棣莫弗最著名的是二项分布的正态逼近. 假如说伯努利大数定律, 是频率在多方面 “平均” 服从某种明显的规律性 (按收敛的形式, 事件出现的频率在一定意义上收敛于其概率), 那么棣莫弗发现的正态逼近, 揭示了关于*偏差*对平均水平性质的另一种广泛的规律性. 棣莫弗的结果及其后来的推广是如此地卓著, 使得人们把 “积分极限定理” 称为概率论的*中心极限定理*.(使用这一术语, 是 G. 波利亚 (George Pólya, 1887 年 — 1985 年) 1920 年建议的, [55].)

在所涉及的时期, 最著名的无疑是拉普拉斯. 他在 1812 年出版的专著《Théorie Analytique des Probabilités》(《概率的分析理论》), 是 19 世纪概率论的主要参考书. 除天文学和数学分析方面的工作之外, 他还撰写了若干关于概率演算的基础, 哲

学问题和具体问题的论文. 拉普拉斯在误差理论方面的贡献卓著. 具体地说, 在误差理论中引进正态律自然的思想属于他和高斯, 正态律是大量独立的、最简单的误差叠加的总效应而产生的结果. 拉普拉斯不仅给棣莫弗积分定理以更加一般的提法 ("棣莫弗 - 拉普拉斯定理"), 并且提出了新的分析证明.

继 J. 伯努利之后, 在有限个可能结局的情形下, 拉普拉斯明确地依照导致概率概念的 "古典型" 定义的、"等可能性原则" 或 "无差异原则".

然而, 在这一时期就已经出现了, 不能置于古典概型中的, "非古典型" 概率分布. 例如, 出现了正态分布和泊松分布. 不过, 这两个分布长时间仅仅被视为某种逼近, 而不是 (像现代对术语的理解) 看成概率分布.

另一个 "非古典型" 分布的例子, 就是 "几何概率的" 问题 (例如, 见牛顿 1665 年, [52] 的 p.60). 这里还有著名的 "蒲丰针" 问题. 与贝叶斯公式相联系, 亦出现了不同的概率. *贝叶斯公式*发表在 1763 年的论文 "An Essay Towards Solving a Problem in the Doctrine of Chances" 中, 在试验出现了某一事件的情况下, 该公式给出了重新计算先验概率的规则 (贝叶斯认为两个概率是相同的)① . 由该公式在统计学中产生了一个完整研究方向, 如今称为 "贝叶斯方法".

综上所述可见, 在概率论的 "古典型" (有限) 框架内本质上制约了其发展和应用的可能性, 而对于正态分布, 泊松分布, 以及其他分布的解释只能局限于某种*极限*形式, 因而产生一种不完善的感觉. 在这个时期, 在概率论中缺乏抽象的数学概念, 而且它还没有被当作应用数学. 况且, 其方法还局限在 (诸如, 赌博, 误差理论, 射击理论, 保险, 人口学 ……) 的具体应用.

3. 第三时期 (19 世纪后半叶) 在这个时期, 彼得堡占据了概率论一般问题的基本位置: П. Л. 切比雪夫 (П. Л. Чебышёв, 亦译 "切贝绍夫", 1821 年 — 1894 年), А. А. 马尔可夫 (А. А. Марков, 1856 年 — 1922 年), 和 А. М. 李雅普诺夫 (А. М. Ляпунов , 1857 年 — 1918 年), 在概率论的整个系统的扩展和深入方面, 作了重要贡献. 具体地说, 由于他们的工作, 突破了 "古典型" 概率的机会的框架. 切比雪夫非常明确地评价了随机变量的概念, 数学期望的概念的作用, 并且有效地演示了这些观念的适用性, 当然这些在现在看来都是很平常的.

大数定律, 棣莫弗 - 拉普拉斯定理涉及仅有两个可能值的随机变量. П. Л. 切比雪夫本质上扩展了这些定理的适用范围 (使之适用于更一般的随机变量). 例如, 他的第一个成果是, 证明了大数定律对于任意独立随机变量之和成立, 只要这些随机变量的绝对值都不大于某一常数. (下一步工作是 А. А. 马尔可夫完成的, 证明用到 "切比雪夫 - 马尔可夫不等式".)

在大数定律之后, П. Л. 切比雪夫转向 "对于独立随机变量之和, 棣莫弗 - 拉普拉斯定理的正确性", 为此他提出了新的证明方法 —— *矩法*, 后来由 А. А. 马尔可夫实现.

① "априори" (a priori) 汉语常译为 "先验", 有的译为 "验前". —— 译者

在寻找棣莫弗 - 拉普拉斯定理成立的一般条件的过程中, A. M. 李雅普诺夫迈出了意想不到的一步, 他用由拉普拉斯开始的特征函数方法证明了, 只要存在 $\delta > 0$ 使被加独立随机变量有 $2 + \delta$ 阶矩, 而没必要存在一切阶矩, 即可证明该定理. 此条件称做 “李雅普诺夫条件”.

作为原则上新的概念应该指出, A. A. 马尔可夫引进的、具有 “无后效” 性的、相依 (非独立) 随机变量, 现在称做 “马尔可夫链”. 并且对于马尔可夫链, 首先严格证明了 “遍历性” 定理.

可以确切地断言, П. Л. 切比雪夫, A. A. 马尔可夫和 A. M. 李雅普诺夫 (“彼得堡学派”) 的工作, 为后来概率论的全部发展奠定了坚实的基础.

19 世纪的后半叶, 在西欧由于发现概率论与纯数学、统计物理的深刻联系, 以及数理统计蓬勃发展, 对于概率论的兴趣开始急速增长.

在这个时期, 概率论本身的发展, 越来越明显地受到其 “古典型” 假设 (结局的有限性及其等可能性) 的强烈制约, 并且需要在纯数学中寻找相应的扩展. (这里应注意, 当时集合论才刚刚建立, 而测度论刚刚处于创立的 “门槛”).

与此同时, 在纯数学中, 特别在数论中, 科学仿佛离概率论十分遥远. 人们开始利用概率的概念, 并且得到了纯 “概率” 本性的结果, 开始对于概率的直观产生兴趣.

例如在 1890 年, J. 庞加莱 (Jules Henri Poincaré, 1854 年 — 1912 年) 在其关于 “三物体” 的论文中得出如下结果: 描绘保持 “体积” 变换 T 的、动态系统运动的返回性的结果说明, 如果 A 是初始状态 ω 的集合, 则对于 “典型的” $\omega \in A$, 运动的轨道 $T^n\omega$ 将无限多次返回集合 A. (按现代语言, 回返不是对于所有初始状态, 而仅对于几乎所有初始状态.)

在这个时期的研究中, 人们常提到 “随机抽样”, “典型情形”, “特殊情形”. 在 J. 庞加莱的教科书《Calcul des Probabilités》([56], 1896 年) 中, 提出一个问题 “在区间 [0,1] 上随机选取的一点, 恰好是有理数的概率如何?”

1888 年天文学家 J. A. H. 盖尔登 (Johan August Hugo Gyldén, 1841 年 — 1896 年) 发表了论文 [18], 论文的起源 (像 J. 庞加莱 [57], 1890 年的论文一样) 与星球的稳定性有关, 而现在应纳入概率数论. 论文的内容如下.

以 “随机地” 选取一个数 $\omega \in [0, 1)$, 并且设 $\omega = (a_1, a_2, \cdots)$ 是 ω 的连续分数分解, 其中 $a_n = a_n(\omega)$ 是整数. (对于有理数 $\omega \in [0, 1)$, 在此分解中仅有有限个 a_n 不为 0, 并且由 $(a_1, a_2, \cdots)$ 形成的数 $\omega^{(k)} = (a_1, a_2, \cdots, a_k, 0, 0, \cdots)$, 用于做 ω 的最优初始逼近.) 问在 “典型” 情形下, 当 n 的值充分大时量 $a_n(\omega)$ 的性质如何?

虽然并不严格, J. A. H. 盖尔登证明, 当 n 充分大时在分解 $\omega = (a_1, a_2, \cdots)$ 中, $a_n = k$ 的值在 “多少” 程度上与 k^2 呈反比. 略晚些, T. 布罗登 (T. Brodén, [12]) 和 A. 威曼 (A. Wiman, [62]) 证明, 利用几何概率, 如果把 “随机” 抽取 $\omega \in [0, 1)$ 看作 ω 在 [0,1) 上 “均匀分布”, 则当 $n \to \infty$ 时 $a_n(\omega) = k$ 的概率收敛于值

$$(\ln 2)^{-1} \times \ln\left[\left(1+\frac{1}{k}\right)\Big/\left(1+\frac{1}{k+1}\right)\right];$$

由此可见, 对于充分大 k 的, 该式与 k^2 呈反比, 而这实质上具有盖尔登的形式.

19 世纪的后半叶, 概率的概念和论断, 开始广泛地应用于经典物理和统计力学. 例如, 只需注意到, 分子运动速度的麦克斯韦分布 (James Clerk Maxwell, 1831 年 — 1879 年; 见 [44]); 布耳兹曼 (Ludwig Boltzmann; 1844 年 — 1906 年) **时间平均值**和**遍历假说** (见 [6], [7]).

总体的概念与他们的名字相联系, 后来总体的概念在吉普斯 (Josiah Willard Gibbs, 1839 年 — 1903 年) 的工作中得到进一步发展 (见 [17]).

对于概率论以后全部的发展, 以及关于对概率方法与概念作用的认识的深化, 如下一些学者都起了重要作用: 1827 年 R. 布朗 (Robert Brown, 1773 年 — 1858 年) 发现的被称为布朗运动的现象 (对该现象的描绘, 1828 年发表在抨击性文章《A Brief Account of Microscopical Observation ⋯》 [11] 中); 在研究铀的性质时, 1896 年 A. 贝克尔雷尔 ((Antoine-) Henri Becquerel, 1852 年 — 1908 年) 发现了**放射性衰变现象**; 1900 年 L. 巴彻里耶 (Louis bachelier, 1870 年 — 1946年, [2]) 利用布朗运动对股票的价格进行数学描绘 (详见 [74]).

A. 爱因斯坦 (Albert Einstein, 1879 年 — 1955 年, [75]) 和 M. 斯莫卢霍夫斯基 (Marian Smoluchowski, 1872 年 — 1917 年, [59]) 后来对布朗运动, 作了定性的解释和定量的描述. 放射性现象在量子力学的框架内得到了说明, 量子力学是 20 世纪的 20 年代创立的.

由以上所述知, 新概率概型和模型的出现, 以及概率的思想体系都超出了 "古典型概率" 的范围, 并且要求有新的概念, 以便对诸如 "来自区间 $[0,1)$ 的样本点" 的含义, 给予确切的数学意义, 更不用说对于 "随机" 布朗运动的解释了. 从这种观点看来, 非常适时的是, 产生了集合论, 和由 E. 博雷尔 (Emile Borel, 1871 年 — 1976 年; [8]) 于 1898 年引进的 "博雷尔测度" 的概念; 以及 H. 勒贝格 (Henri Lebesgue, 1875 年 — 1941 年) 积分理论, 包含在其 1904 年的书中 [40]; 由 E. 博雷尔作为长度概念的推广, 将测度引进欧几里得空间. 遵循 M. 弗雷歇 (Maurice Fréchet, 1878 年 — 1973 年; 1915 年 [71]) 现代讲述测度论, 是在**抽象**可测空间上 (关于测度论和积分的历史, 例如见 [72]).

实际上, 立即可以察觉, 博雷尔测度论和勒贝格积分理论概念的基础, 不但可以为许多研究奠定基础, 而且为诸如 "在区间 $[0,1)$ 是随机选点" 之类的许多直观提法赋予确切的意义. 而且, E. 博雷尔本人很快 (1905 年) 就将理论 - 集合方法应用于概率论, 证明了 —— 强大数定律 —— 关于实数的某些性质, "以概率 1" 成立或 "几乎必然" 成立.

这些定理给出一定的印象, "或多或少" 的实数, 具有 (在如下指出意义上) "特殊" 性质. 这些定理的实质如下.

假设实数 $\omega \in [0,1)$, 而 $\omega = 0.\alpha_1\alpha_2\cdots$ 是二进制分解, 其中 $\alpha_n = 0,1$ (对照上面的 ω 分解为连续分数 $\omega = (\alpha_1, \alpha_2, \cdots)$). 那么, 若 $v_n(\omega)$ 是在 ω 的前 n 个值 $\alpha_1, \cdots, \alpha_n$ 中 "1" 的个数, 则满足条件: 当 $n \to \infty$ 时 $v_n(\omega) \to 1/2$ 的 ω (博雷尔称之为 "正规的") 集合的博雷尔测度等于 1, 而对于那些 ("特殊的") $\omega, v_n(\omega)$ 不收敛, 且相应集合的博雷尔测度等于 0.

这一结果 ("博雷尔强大数定律") 外观上像 J. 伯努利定理 ("大数定律"). 然而, 在二者之间, 既有形式上数学的不同, 又有哲学概念上的差异. 事实上, 在大数定律中仅仅断定: 对于任意 $\varepsilon > 0$, 当 $n \to \infty$ 时事件 $\{\omega : |v_n(\omega) - 1/2| \geqslant \varepsilon\}$ 的概率收敛于 0. 而在强大数定律中结果更多: 事件

$$\left\{\omega : \sup_{m \geqslant n} |v_m(\omega) - 1/2| \geqslant \varepsilon\right\}$$

的概率趋向 0. 此外, 在第一种情形下, 命题涉及有限序列 $(\alpha_1, \alpha_2, \cdots, \alpha_n), n \geqslant 1$ 概率的某些性质, 以及这些概率的极限. 而在第二种情形下, 命题涉及定义在无限序列 $(\alpha_1, \alpha_2, \cdots, \alpha_n, \cdots)$ 概率的性质, 以及这些概率的极限. (与概率论方法渗透到数论中, 以及构建现代概率论, 有关的数学和哲学问题的广泛资料详细阐述, 参见专著 "Creating Modern Probability", [54], 作者是 J. v. 普拉托 (Jan von Plato)).

4. 第四时期 (20 世纪初和 20 世纪中叶) 19 世纪末, 表现出概率论与纯数学的联系, 使得 D. 希尔伯特 (David Hilbert, 1862 年 — 1943 年) 提出*概率论数学化的问题*. 1900 年 8 月 8 日, 在巴黎召开的第二届数学学术会议上, D. 希尔伯特在他的提纲性报告中提出上述问题. 在他的著名课题中 (第一个是关于连续统 - 假说), 第六个是数学起决定性作用的物理学科的公理化问题. D. 希尔伯特把概率论和力学归为这样的学科, 他还指出在物理学, 其中包括气体动能理论, 严格的和尚未满意的发展平均值方法的必要性. D. 希尔伯特指出, 格丁根 (Göttingen) 大学①的 G. 布耳曼 (Georg BoHlmann, 1869 年 — 1928 年) 副教授, 由他提出了 "关于概率论的公理化问题的动议", 他 1900 年在巴黎召开的保险统计人员会议上作了关于公理化问题的发言 (见 [5], [19]). G. 布耳曼引进的概率定义为事件上的 (有限 - 可加) 函数, 然而没有 "事件系" 的充分清晰的定义, 其实他自己也承认这一点.

概率论形成历史的第四个时期 —— 是概率论的逻辑奠基和形成数学学科的时期.

在 D. 希尔伯特的报告之后, 很快出现了几个建立概率的数学理论的尝试, 其共同的特点是, 以集合论与测度论为基础.

例如, 1904 年 R. 拉默尔 (R. Lämmel, [41]; 亦见 [19]) 为描绘结局的集合, 他试图利用集合论, 不过, 概率的概念 (使用术语 "content", 且结合体积, 面积, 长度 ……)

①格丁根是德国城市, 人口十几万. 格丁根大学是一所综合性大学 1737 年创立, 也是该城唯一一所高校, 学生两万人左右. 格丁根科学院 1751 年成立, 有近百名院士和百余名通讯院士. —— 译者

本身仍然停留在以往时期的直观水平上.

另一个作者 U. 布罗吉 (Ugo Broggi, 1880 年 — 1965 年) 在其由 D. 希尔伯特指导的学位论文 (1907 年, [10]; 亦见 [19]) 中, 也运用博雷尔和勒贝格测度论 (基于它在 1904 年勒贝格的书 [40]) 中的概念, 但是 (有限 - 可加) 概率的概念本身 (在最简单的场合) 需要运用 "相对测度", "相对频率" 和 (在一般情形下) 某种人为的极限过程.

伯恩斯坦 在随后关于概率论的逻辑奠基工作作者中, 首先应当提到的是, C. H. 伯恩斯坦 (C. H. Бернштейн, 1880 年 — 1968 年) 和 R. von 米泽斯 (Richard von Mises, 1883 年 — 1953 年).

C. H. 伯恩斯坦的公理化体系 ([4], 1917 年), 以事件按其大或小的似然程度的质量比较为基础, 而概率的数值本身表现为某个任意值.

费内提 后来, 在 B. de 费内提 (Bruno de Finetti, 1906 年 — 1985 年) 20 年代末 — 30 年代初的工作, 用基于与主观数量判断十分相似的方法 ("主观知识体系"), 得到广泛的发展 (例如, 见 [65]~[70]).

B. de 费内提的思想, 得到统计学中许多贝叶斯方向代表的很大支持, 例如, L. J. 塞维奇 (Leonard Jimmie Savage, 1917 年 — 1971 年; [60]). B. de 费内提的思想, 在有相当大的作用的对策论和判决理论中, 也得到同样的支持.

米泽斯 1919 年 R. 米泽斯提出了([49], [50]), 论证概率论的所谓*频率方法* (亦称*统计方法*或*经验方法*). 他的所谓*频率方法*以如下的思想为基础, 概率的概念只可能用于所谓 "集体", 即*个别无限有序数列*, 且这种有序数列具有其形成的某种 "随机" 性质.

R. 米泽斯的一般概形描绘如下.

考虑 "试验" 结局的某个样本空间, 并且假设可以进行无限多次试验, 得序列 $x = (x_1, x_2, \cdots)$, 其中 x_n 是第 n 次 "试验" 的结局. 其次, 设 A 是试验结局的集合的子集, 而

$$v_n(A; x) = \frac{1}{n}\sum_{i=1}^{n} I_A(x_i)$$

是 "事件" A 在前 n 次 "试验" 中出现的频率.

序列 $x = (x_1, x_2, \cdots)$ 称做*集体*, 如果它满足如下 (称为*米泽斯泽一性条件* (见 [49]~ [51]) 的) 两个假说.

I (对于*序列频率的极限存在*) 对于一切 "容许" 集合 A, 频率的极限存在:

$$\lim_n v_n(A; x)\ (= p(A; x));$$

Ⅱ (对于*子序列频率的极限存在*) 对于一切由序列 $x = (x_1, x_2, \cdots)$, 利用某一事先约定的 ("容许") 其形成规则体系 (米泽斯称之为*位选择函数* [Place–selection functions]), 得到的子序列 $x' = (x'_1, x'_2, \cdots)$, 频率 $\lim_n v_n(A; x')$ 的极限与关于序列 $x = (x_1, x_2, \cdots)$ 本身的极限一样, 即等于 $\lim_n v_n(A; x)$.

按照米泽斯, 只有联系具体的“集体”才可以称“集合 A 的概率”, 而且 (根据假说 I) 这一概率 $(P(A;x))$ 定义为频率 $\lim_n v_n(A;x)$的极限. 需要强调, 假如该极限不存在 (即按照定义 x, 不是“集体”), 那么相应的概率也就没有定义. 第二个假说, 米泽斯用来表示 (相应的直观, 恰好是一切概率研究的基础) 在形成“集体” $x=(x_1,x_2,\cdots)$ 时“随机性”概念, 反映该序列“非正则性”思想, 以及对于任意 $n\geqslant 1$, 根据“过去” $(x_1,x_2,\cdots,x_{n-1})$, 其“将来”值 $(x_n,x_{n+1},\cdots)$ 的“不可预测性”. (赞同第二章 §1 介绍的, 柯尔莫戈洛夫公理化体系的, 概率论的代表人物, 这样的序列应当联想到, 对独立同分布随机变量观测结局的“典型”序列; 见第一章 §5 第 4 小节.)

米泽斯, 像他自己 ([51] 第 1 页) 说的那样, 在建立 “a mathematical theory of repetitive events” (重复事件的数学理论) 时提出的假说, (特别是在 30 年代) 引起了很大的争议和批评. 基本的不同意见如下: 实际中人们遇到的一般是有限序列, 而不是无限序列. 因而, 实际上无法确定极限 $\lim_n v_n(A;x)$ 是否存在; 在从序列 x 转移到序列 x' 时, 实际上也无法确定该极限《敏感性》. 米泽斯定义的子序列形成“容许”规则的概念, 也受到严重的批评, 使得在非此即彼的条件 II 中, 出现许多 (“模棱两可”) 规则定义的模糊不清晰性.

如果考虑由 0 和 1 形成的序列 $x=(x_1,x_2,\cdots)$, 而且对于该序列, 极限 $\lim_n v_n(x;\{1\})$ 的值属于区间 (0,1), 则该序列既含无限个 0 又含无限个 1. 因此, 如果容许任何形成子序列的规则, 那么, 例如可以由 x 组成仅含“1”的子序列 x', 显然 $\lim_n v_n(x';\{1\})=1$. 由此可见, 关于一切形成子序列的方法的非平凡“集体”不存在.

在证明“集体”类“非空”的第一步, 是由 A. 瓦尔德 (Abraham Wald, 1902 年 — 1950 年) 1937 年在其论文 [13] 中实现的. 在他由序列 $x=(x_1,x_2,\cdots)$ 构造子序列 $x'=(x'_1,x'_2,\cdots)$ 时, 利用只取 0 和 1 两个可能值的可数个函数组 $f_i=f_i(x_1,\cdots,x_i), i\geqslant 1$: 若 $f_i(x_1,\cdots,x_i)=1$, 则元素 x_{n+1} 属于子序列 x'; 若 $f_i(x_1,\cdots,x_i)=0$, 则元素 $x_{n+1}\notin x'$. 1940 年 A. 乔尔奇 (Alonzo Church, 1903 年 — 1995 年) 提出形成子序列的另一种方法 (见 [73]), 方法基于如下想法: 这样的形成方法应该是实际上“可有效计算的”. 乔尔奇的这一思想导致, 他为序列的建立而提出的, 函数计算算法的概念. (例如, 设 x_i 有两个可能值: $\omega_1=0,\omega_2=1$. 将序列与一正数

$$\lambda=\sum_{k=1}^{n} i_k 2^{k-1}$$

相对应, 其中决定于 $i_k=\omega_{i_k}$. 如果 $\varphi=\varphi(\lambda)$ 是定义在集合 $\{0,1,2,\cdots\}$ 上的, 给定的二进制 $\{0,1\}$ - 函数, 则当 $\varphi(\lambda_n)=1$ 时 x_{n+1} 包含新序列 x' 中, 而当 $\varphi(\lambda_n)=0$ 时 x_{n+1} 不包含 x' 中).

“集体”作为具有“随机性”序列的说明和证据之一, 米泽斯引进了直观的论据: 对于这样的序列不能构造“能获胜的对策体系”.

在 J. 维尔 (Jean Ville, 1910 年 — 1988 年) 1939 年不大的专著 [14] 中, 对这样的论断进行了批评性分析, J. 维尔在 [14] 中给米泽斯的结果以严格的数学形式. 需

要指出, 也正是在此专著中, 首次 (作为数学概念) 使用术语 “鞅”.

由以上引进的概率论公理化不同方法 (······, 伯恩斯坦, de 费内提, 米泽斯) 的描述可见, 在他们那里留有概念的复杂化与过分超负荷错误的印记, 其中多是出于建立尽可能接近实际应用的、概率概形的愿望. 正像 A. H. 柯尔莫戈洛夫在其专著《概率论的基本概念》[23] 中指出的, 这无法产生简单的公理化体系.

A. H. 柯尔莫戈洛夫, 提到他对于概率论逻辑基础的兴趣的, 发表的第一篇作品, 是 (非广为已知的) 论文 “测度与概率演算的一般理论” [27]. 论文的名称及其内容表明, A. H. 柯尔莫戈洛夫把集合论与测度论, 看作概率论逻辑基础的可能性所在. 由以上的叙述可见, 这种情况完全不是新的, 并且对于莫斯科数学学派是完全自然的. 因为对于莫斯科数学学派, 集合论及度量函数理论, 是数学研究的基本领域之一.

在这篇论文 (1929 年) 与《概率论的基本概念》([23], 1933 年) 之间, A. H. 柯尔莫戈洛夫发表了他著名的论文之一 “概率论的分析方法” [29]. 关于这篇论文, П.С. 亚历山大罗夫 (П. С. Александров) 和辛钦 [1] 写道: “在整个 20 世纪的整个概率论中, 很难指出对于未来科学的发展如此奠基性的研究成果 ······”.

这篇论文的重大价值在于, 它不仅奠定了*马尔可夫随机过程*的理论基础, 而且它表明概率论整体上与数学分析 (特别是与常微分和偏微分方程论) 的联系, 以及与经典力学和经典物理学等的联系.

鉴于所考虑的概率的数学理论基础问题, 我们指出, 譬如说论文 “分析方法” [29], 可以看作逻辑上建立随机过程基础必要性的、“物理的” 动机综合体, 是 (除 “公理化” 外) “基本概念” 链中的一环.

A. H. 柯尔莫戈洛夫提出的 (非正式的) 概率论的公理化, 基础是*概率空间*

$$(\Omega, \mathscr{F}, \mathbf{P})$$

的概念, 其中 $(\Omega, \mathscr{F})$ 是 (“基本” 结局和 “事件” 的) 某一 (抽象) 可测空间, 而 $\mathbf{P}$ 是 $\mathscr{F}$ 中的非负可数 - 可加集函数, 且满足规范性条件 $\mathbf{P}(\Omega) = 1$ (“概率”); 见第二章 §1.

关于*随机变量* $\xi = \xi(\omega)$ 理解为 $\mathscr{F}$ - 可测函数 $\xi = \xi(\omega)$, 而按测度 $\mathbf{P}$ 定义的 $\xi(\omega)$ 的勒贝格积分是其*数学期望*.

关于 $\mathscr{G} \subseteq \mathscr{F}$ 的 σ– 子代数*条件数学期望* $\mathbf{E}(\xi|\mathscr{G})$ 是新概念 (见 A. H. 柯尔莫戈洛夫的《概率论的基本概念》(第二版) 的前言 [24]).

在《概率论的基本概念》中, 有一定理, A. H. 柯尔莫戈洛夫称之为*基本的*, 这本身就强调它所包含 (关于具有给定有限维分布的存在性的过程) 论断的重要性. 这里, 问题的实质如下.

在 A. H. 柯尔莫戈洛夫的论文 “概率论的分析方法” [29] 中, 马尔可夫过程用于 “随机确定系统” 的发展变化, 这用满足 “柯尔莫戈洛夫 - 查普曼方程” 的、函数 $P(s, x; t, A)$ 性质的语言描述. 函数 $P(s, x; t, A)$ 称做转移概率函数, 因为它表示 “在时间 s ‘系统’ 处于状态 x, 而在时间 t ‘系统’ 处于其状态相空间的集合 A 中”.

同样, 同一时间, 在 [30], [31], [66], [67] 关于 “齐次独立增量随机过程” 的工作中, 所有的研究都用满足函数方程

$$P_{s+t}(x)=\int P_s(x-y)\mathrm{d}P_t(y)$$

的函数 $P_t(x)$ 性质的语言进行的, 在说明 $P_t(x)$ 是 “经时间 t 过程的增量不大于 x 的概率” 时, 上面的方程是自然产生的.

不过, 从形式逻辑的观点看, 可以称为 “具有给定转移概率 $P(s,x;t,A)$ 或给定分布 $P_t(x)$ 的 ‘过程’ ” 的对象之*存在性*问题, 仍然没有解决.

就是在求解该问题时, 涉及如下*基本定理*: 对于每个一致有限维概率分布组

$$F_{t_1,t_2,\cdots,t_n}(x_1,x_2,\cdots,x_n),\quad 0\leqslant t_1<t_2<\cdots<t_n, x_i\in\mathbb{R},$$

可以建立一个概率空间 $(\Omega,\mathscr{F},\mathbf{P})$ 和随机变量组 $X=(X_t)_{t\geqslant 0}, X_t=X_t(\omega)$, 使得

$$\mathbf{P}\{X_{t_1}\leqslant x_1,X_{t_2}\leqslant x_2,\cdots,X_{t_n}\leqslant x_n\}=F_{t_1,t_2,\cdots,t_n}(x_1,x_2,\cdots,x_n).$$

作为 Ω 取实函数 $\omega=(\omega_t)_{t\geqslant 0}$ 的空间 $\mathbb{R}^{[0,\infty)}$, 作为 $\mathscr{F}$ 取柱集生成的最小 σ - 代数, 而测度 $\mathbf{P}$ 是由柱集代数上的测度 (在柱集代数上, 该测度是自然地根据有限维概率分布建立的), 开拓到 $\mathscr{F}$ 上概率测度. 随机变量组 $X_t(\omega)$ 是用坐标方式建立的: 若 $\omega=(\omega_t)_{t\geqslant 0}$, 则 $X_t(\omega)=\omega_t$. (这一构造说明了, 为何 “随机过程” 的概念, 常等同于 (它) 在函数空间 $\mathbb{R}^{[0,\infty)}$ 中的*测度*).

在《概率论的基本概念》中, 以不大的篇幅讲概率论的*适用性*问题.

在描绘将概率论用于 “现实试验世界” 的*条件模式*时, A. H. 柯尔莫戈洛夫许多方面是按米泽斯做的, 而且表明在概率论的解释与应用性问题上它与米泽斯方法并不是格格不入的.

关于这一*条件模式*的实质可以描述如下.

假设有某一条件的*综合体*, 使得可以进行无限次重复试验.

设 $(x_1,x_2,\cdots,x_n)$ 表示 n 次试验的结果, 其中例如可以假设 $x_i(1\leqslant i\leqslant n)$ 属于集合 X. 此外, 设 A 是我们感兴趣的 X 的某一子集.

如果 $x_i\in A$, 则称第 i 次试验出现了事件 A. (注意, *事先*不作诸如: 试验是 “随机和独立地进行的” 等, 关于试验的 “概率” 性质之类的任何假设; 对于导致事件 A 的 “情形” 也不作任何假设, ……)

其次, 假设可以赋予事件 A 某个数 (记作 $\mathbf{P}(A)$), 使得*实际上可以认为*, 在 n 次试验中事件 A 出现的频率 $v_n(A)$, 对于充分大的 n, 与 $\mathbf{P}(A)$ 的差异很小. 而如果 $\mathbf{P}(A)$ 较小, 那么*实际上可以认为*, 在一次试验中事件 A 不出现.

在《概率论的基本概念》中, A. H. 柯尔莫戈洛夫, 在未讨论概率论对于《现实世界》可应用性条件细节的情况下, 写道 “我们 …… 有意识地把经验世界中, 关于

概率概念的高深哲学著作放在一边", 但是在第一章引言中指出, 存在概率论的应用领域, 其中 "与 '偶然性' 和 '概率' 概念等词的本意毫无关系" (见 [24]).

三十年后, A. H. 柯尔莫戈洛夫又回到概率的可应用性问题 (见 [32]~[37]), 对于该问题的解, 他提出两种 ("第一种" 和 "第二种") 处理方法, 相应为 "公理化随机性" 的概念和 "算法的复杂性" 的概念. 他这时特别强调 [37], 与定义了*无限序列* $(x_1, x_2, \cdots)$ 的 P. 米泽斯和 A. 乔尔奇不同, 他对于 "随机性" 概念的处理方法, 带有*严格有限性*, 即涉及*有限*长度序列 $(x_1, x_2, \cdots, x_N), N \geqslant 1$ (下面 [38] 的情形称为链), 这实际上是现实中的情形.

"公理化随机性" 是这样引进的.

设 $(x_1, x_2, \cdots, x_N)$ 是以长为 $N(n \leqslant N)$ 的二进制 $x_i = 0, 1$ 链. 称此链关于 (有限) *容许算法*的全体 Φ 是 (n, ε) - *随机的*, 如果存在这样一个数 $p(= \mathbf{P}(\{1\}))$, 使对于利用某个算法 $A \in \Phi$, 由 $(x_1, x_2, \cdots, x_N)$ 得到的任何链 $(x'_1, x'_2, \cdots, x'_m), n \leqslant m \leqslant N$, 出现 "1" 的频率与 $v_m(x'; \{1\})$ 离 p 的差异不大于 ε. (算法的全体 Φ 中导致长为 $m < n$ 的链, 不予考虑).

A. H. 柯尔莫戈洛夫在 [32] 中证明, 假如对于给定的 n 和 $0 < \varepsilon < 1$, 容许算法的个数不大于

$$\frac{1}{2} \exp\{2n\varepsilon^2(1-\varepsilon)\},$$

则对于每一个 $0 < p < 1$和任意 $N \geqslant n$, 存在具有 (n, ε) - 随机性 ("公理化随机性") 的链 $(x_1, x_2, \cdots, x_N)$.

由于容许描绘和选择算法的不确定性, 在上面描绘的分出 "随机" 链 (像米泽斯情形一样) 的方法, 有一定任意性. 这时这一算法类明显不会太大, 否则 "公理化随机" 链的集合成为空集. 与此同时, 总是希望容许算法类构造简单 (例如可以用表格表示).

在概率论中形成一种完全确定、加强不同类型的、关于 "典型随机实现相当不正规的, 十分复杂的" 概率论点的观念.

因此, 如果倾向于, 使链及序列 "随机性" 的*算法*定义, 最大限度地接近关于随机实现构造的*概率*概念, 则 Φ 中的算法 (的全体) 容许挑出 "非典型的, 容易剔除的" 链, 把那些充分非正则的, 十分复杂的定为 "随机的".

这种理由导致 A. H. 柯尔莫戈洛夫的 "第二种" 处理方法, 导致 "随机性" 概念, 其中着重点不是所考虑的算法 "简单", 而是链本身的 "复杂性" 和 "径直地" 引进 "复杂性的" 某个数字特征, 使之表示形成该链时 "非正则性" 程度.

这一数字特征就是单个链 x 关于算法 A 的所谓 "*算法的*" (或 "*柯尔莫戈洛夫的*") *复杂性* $K_A(x)$; 形象地说, $K_A(x)$ 定义为满足如下条件的二进制链的长度: 在算法 (机器, 计算机 ……) 的 "入口" 输入 A, 容许同一链在 "出口" 再现.

正式的定义如下.

设 $\sum$ 是一切有限二进制链 $(x_1, x_2, \cdots, x_n)$ 的全体, 以 $|x|(= n)$ 表示链的长度,

而 Φ 是某一算法类. 实数

$$K_A(x) = \min\{|\rho| : A(\rho) = x\},$$

称做链 $x \in \sum$ 关于算法 $A \in \Phi$ 的复杂性. 数 $K_A(x)$ 是在算法 A 的 "入口" 的二进制链 ρ 的最小长度 ($|\rho|$), 并且在链的 "出口" $x(A(\rho) = x)$ 将得到恢复.

A. H. 柯尔莫戈洛夫在 [34] 中证明, (对于某一重要算法类 Φ) 有如下结果: 存在这样的通用算法 $U \in \Phi$, 对于任意 $A \in \Phi$ 存在常数 $C(A)$, 使对于任意链 $x \in \sum$, 有

$$K_U(x) \leqslant K_A(x) + C(A),$$

而对于通用算法 U' 和 U'',

$$|K_{U'}(x) - K_{U''}(x)| \leqslant C, \quad x \in \sum,$$

其中 C 不依赖于 $x \in \sum$. (A. H. 柯尔莫戈洛夫在 [34] 中指出, P. 所罗门诺夫 [P. Соломонов] 同时证明了类似的结果).

这一事实 (以及对于 "典型" 链 $x, K_U(x)$ 的 $|x|$ 值随增长而增长) 决定如下定义: 量 $K(x) = K_U(x)$ 称做链 $x \in \sum$ 关于算法类 Φ 的复杂性, 其中 U 是 Φ 中的某一通用算法.

量 $K(x)$ 通常称做 "对象" x 算法复杂性或柯尔莫戈洛夫复杂性. A. H. 柯尔莫戈洛夫把该量看作包含在 "有限对象" x 中的算法信息质量的度量, 并且称之为 x 的熵. 并且认为这个概念是比概率信息量的概念更基本, 而为概率信息量的要求知道在 "对象" x 上的概率分布.

量 $K(x)$ 还可以视为 "文本" x 的压缩程度的指标. 如果在类 Φ 中包含元素的简单计数算法, 则很明显 (精确到常数) 链 x 的 "复杂性" $K(x)$ 不大于其长度 $|x|$. 另一方面, 由简单的讨论可见, "复杂性" 小于 K 的 (二进制) 链 x 的数量不大于 $2^K - 1$, 等于小于长度 K 的不同 "输入" 二进制序列的个数:

$$1 + 2 + \cdots + 2^{K-1} = 2^K - 1.$$

其次, 通过简单的论证 (例如, 见 [15]) 可以证明, 存在这样的链 x: 其 "复杂性" (精确到常数) 等于 (且不可能大于)长度 $|x|$, 并且容许强压缩 ("复杂性" 为 $n - a$ 的链的比率不大于 2^{-a}). 由全部所作的这些论述, 自然地导致如下的定义: 算法的 "复杂性" $K(x)$ 接近 $|x|$ 的链 x 称为 (关于算法类 Φ) "算法上随机的".

换句话说, 算法的处理方法①判定这样一些链 x 为 "随机的", 假如其 "复杂性" 是最大的 ($K(x) \sim |x|$).

A. H. 柯尔莫戈洛夫引进的 "复杂性" 的概念, 产生了整个算法随机性的方向, 称为 "柯尔莫戈洛夫复杂性", 在最广泛的数学及其应用领域得到众多的应用 (例如, 详见 [38], [45]~[48], [22]).

①即前面提到的, A. H. 柯尔莫戈洛夫 "第二种" 处理方法, 对应于 "算法的复杂性". —— 译者

在概率论中, 这些新概念是一系列工作的开端: 说明对于何种链和序列的“算法随机性”, 概率—统计规律性 (诸如, 强大数定律, 重对数定律 ……) 成立 (例如, 见 [16]), 从而有可能运用概率论的方法及其结果, 而在前面已经指出的那些领域 (见 [24]), “与‘偶然性’和‘概率’概念等词的本意毫无关系”.



“概率的数学理论形成的简史”的参考文献

[1] Александров П. С., Хинчин А. Я. Андрей Николаевич Колмогоров (к пятидесятилетию со дня рождения) // Успехи математических наук. — 1953. — Т. 8. №3. — С. 177 — 200.

[2] Башелье (Bachelier L.). Théorie de la spéculation // Annales de l'École Normale Supérieure. — 1900. — V. 17. — P. 21 — 86.

[3] Бернулли Я. О законе больших чисел. Ч.4: Искусство предположений. — М.: Наука, 1986. — С. 23—59.

[4] Бернштейн С. Н. Опыт аксиоматического обоснования теории вероятностей // Сообщения Харьковского математического общества. Сер. 2. — 1917. — Т. 15. — С. 209 — 274.

[5] Больман (Bohlmann G.). Lebensversicherungsmathematik // Encyklopaedieder mathematischen Wissenschaften. — Bd. I, Heft 2. — Artikel ID4b. — Leipzig: Teubner, 1903.

[6] Больцман (Boltzmann L.). Wissenschaftliche Abhandlungen. — V.1 — 3. — Leipzig: Barth, 1909.

[7] Больцман, Набл (Boltzmann L., Nabl J.). Kinetische Theorie der Materie // Encyklopaedie der mathematischen Wissenschaften. — Bd. V, Heft 4. — Leipzig: Teubner, 1907. — S. 493 — 557.

[8] Борель (Borel É.). Leçons sur la théorie des fonctions. — Paris: Gauthier–Villars, 1898; Éd. 2. — Paris: Gauthier–Villars, 1914.

[9] Борель (Borel É.). Quelques remarques sur les principes de la theorie desensembles // Mathematische Annalen. — 1905. — V. 60. — P. 194 — 195.

[10] Брогги(Broggi U.). Die Axiome der Wahrscheinlichkeitsrechnung. Dissertation. Göttingen, 1907. (См. также [19].)

[11] Браун (Brown R.). A brief account of microscopical observations made in the months of June, July, and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. — Philosophical Magazine N. S. — 1828. V. 4. — P. 161 — 173.

[12] Броден (Brodén T.). Wahrscheinlichkeitsbestimmungen bei der gewöhnlichen Kettenbruchentwicklung reeller Zahlen // Akad. Förh. Stockholm. — 1900. — V. 57. — P. 239 — 266.

[13] Вальд(Wald A.). Die Widerspruchsfreiheit des Kollektivbegriffes der Wahrscheinlichkeitsrechnung // Ergebnisse eines mathematischen Kolloquiums. —1937. — V. 8. — P. 38 — 72.

[14] Виль (Ville J. A.). Étude critique de la notion de collectif. — Paris: Gauthier–Villars, 1939.

[15] Витаньи П., Ли М. Колмогоровская сложность: двадцать лет спустя // Успехи математических наук. — 1988. — Т. 43, № 6. — С. 129 — 166.

[16] Вовк В. Г. Закон повторного логарифма для случайных по Колмогорову, или хаотических, последовательностей // Теория вероятностей и ее применения. — 1987. — Т. 32, № 3. — С. 456 — 468.

[17] Гиббс (Gibbs J. W.). Elementary Principles in Statistical Mechanics. Developed with especial reference to the rational foundation of thermodynamics. — New Haven: Yale Univ. Press, 1902; New York: Dover, 1960.

[18] Гюлден (Gyldén H.). Quelques remarques relativement à la représentation de nombres irrationnels au moyen des fractions continues // Comptes Rendus. Paris. — 1888. — V. 107. — P. 1584 — 1587.

[19] Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933 / Ed. I. Schneider. — Berlin: Akademie–Verlag, 1989.

[20] Дэвид (David F. N.). Games, Gods and Gambling. The Origin and History of Probability and Statistical Ideas from the Earliest Times to the Newtonian Era. — London: Griffin, 1962.

[21] Звонкин А. К., Левин Л. А. Сложность конечных объектов и обоснование понятий информации и случайности с помощью теории алгоритмов // Успехи математических наук. — 1970. — Т. 25, № 6. — С. 85 — 127.

[22] Кирхгер, Ли, Витаньи (Kircherr W., Li M., Vitányi P.). The miraculous universal distribution // Mathematical Intelligencer. — 1997. — V. 19, № 4. — P. 7 — 15.

[23] Колмогоров (Kolmogoroff A.). Grundbegriffe der Wahrscheinlichkeitsrechnung. — Berlin: Springer, 1933; Berlin — New York: Springer, 1973.

[24] Колмогоров А. Н. Основные понятия теории вероятностей. — М. — Л.: ОНТИ, 1936; 2-е изд. — М.: Наука, 1974; 3-е изд. — М.: ФАЗИС, 1998. 汉译本: 《概率论的基本概念》(丁寿田译), 商务印书馆, 1952.

[25] Колмогоров (Kolmogorov A. N.). Foundations of the Theory of Probability. — New York: Chelsea, 1950; 2nd ed. — New York: Chelsea, 1956.

[26] Колмогоров А. Н. Роль русской науки в развитии теории вероятностей // Роль русской науки в развитии мировой науки и культуры. — Т. I, кн. 1. — М.: Изд-во Моск. ун-та, 1947. — С. 53 — 64.

[27] Колмогоров А. Н. Общая теория меры и исчисление вероятностей // Коммунистическая академия. Секция естественных и точных наук. Сборник работ математического раздела. — Т. 1. — М., 1929. — С. 8 — 21. (См. также [28], с. 48 — 58.)

[28] Колмогоров А. Н. Теория вероятностей и математическая статистика. М.: Наука, 1986.

[29] Колмогоров (Kolmogoroff A.). Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung // Mathematische Annalen. — 1931. — V. 104. — P. 415 — 458. (См. также [28], с. 60 — 105.)

[30] Колмогоров (Kolmogoroff A.). Sulla forma generale di un processo stochastico omogeneo. (Un problema di Bruno de Finetti.) // Atti della Accademia Nazionale dei Lincei. — 1932. — V. 15. — P. 805 — 808. (См. также [28].)

[31] Колмогоров (Kolmogoroff A.). Ancora sulla forma generale di un processo omogeneo // Atti della Accademia Nazionale dei Lincei. — 1932. — V. 15. — P. 866 — 869. (См. также [28].)

[32] Колмогоров (Kolmogorov A. N.). On tables of random numbers // Sankhyā A. — 1963 — V.25, № 4. — P. 369 — 376.

[33] Колмогоров А. Н. Теория информации и теория алгоритмов. М.: Наука, 1987.

[34] Колмогоров А. Н. Три подхода к определению понятия 《количество информации》 // Проблемы передачи информации. — 1965. — Т. 1, №1. — С. 3 — 11. (См. также [33], с. 213 — 223.)

[35] Колмогоров (Kolmogorov A. N.). Logical basis for information theory and probability theory // IEEE Transactions on Information Theory. — 1968. — V. 14, № 5. — P. 662 — 664. (См. также [33], с. 232 — 237.)

[36] Колмогоров А. Н. Комбинаторные основания теории информации и исчисления вероятностей // Успехи математических наук. — 1983. — Т. 38, № 4. — С. 27 — 36.

[37] Колмогоров(Kolmogorov A. N.). On logical foundations of probability theory // Probability Theory and Mathematical Statistics (Tbilisi, 1982). Berlin etc.: Springer-Verlag, 1983. — P. 1 — 5. — (Lecture Notes in Mathematics; V. 1021) (См. также [28], с. 467 — 471.)

[38] Колмогоров А. Н., Успенский В. А. Алгоритмы и случайность // Теория вероятностей и ее применения. — 1987 — Т. 32, № 3. — С. 425 — 455.

[39] Лаплас (Laplace P. S., de). A Philosophical Essay on Probabilities. — New York: Dover, 1951; — Первое издание: La Place P. S. Essai philosophique surles probabilités. — Paris, 1814.

[40] Лебег (Lebesgue H.). Leçons sur l'intégration et la recherche des fonctions primitives. — Paris: Gauthier-Villars, 1904.

[41] Леммель (Lämmel R.). Untersuchungen über die ermittlung der Wahrscheinlichkeiten. Dissertation. Zürich, 1904. (См. также [19].)

[42] Ли, Витаньи (Li M., Vitányi P. M. B.). An Introduction to Kolmogorov Complexity and its Applications. — 2nd ed. — Berlin — New York: Springer-Verlag, 1997.

[43] Майстров Л. Е. Теория вероятностей. Исторический очерк. — М.: Наука, 1967.

[44] Максвелл (Maxwell J. C.). The scientific letters and papers of James Clerk Maxwell. — V. I: 1846 — 1862. — V. II: 1862 —1873. — V. III: 1874 – 1879 / Ed. P. M. Harman. — Cambridge: Cambridge Univ. Press, 1990, 1995, 2002.

[45] Мартин-Лёф П. О понятии случайной последовательности // Теорияв-ероятностей и ее применения. — 1966. — Т. 11. № 1. — С. 198 — 200.

[46] Мартин-Лёф (Martin-Löf P.). The definition of random sequences // Information and Control. — 1966. — V. 9, № 6. — P. 602 — 619.

[47] Мартин-Лёф (Martin-Löf P.). On the notion of randomness // Intuitionism and Proof Theory: Proceedings of the conference at Buffalo, NY, 1968 / Ed. A. Kino et al. Amsterdam: North–Holland, 1970. — P.73 — 78.

[48] Мартин-Лёф (Martin-Löf P.). Complexity oscillations in infinite binary sequences // Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete. — 1971. — V. 19. — P. 225 — 230.

[49] фон Мизес (Mises R., von). Fundamentalsätze der Wahrscheinlichkeitsrechnung // Mathematische Zeitschrift. — 1919. — V. 4. — P. 1 — 97.

[50] фон Мизес (Mises R., von). Grundlagen der Wahrscheinlichkeitsrechnung // Mathematische Zeitschrift. — 1919. — V. 5. — P. 52 — 99; 1920. — V. 7. — P. 323.

[51] фон Мизес (Mises R., von). Mathematical Theory of Probability and Statistics. — New York — London: Academic Press, 1964.

[52] Ньютон (Newton I.). The Mathematical Works of Isaac Newton / Ed. D. T. Whiteside. — V. 1. — New York: Johnson, 1967.

[53] О теории вероятностей и математической статистике (переписка А. А. Маркова и А. А. Чупрова). — М.: Наука, 1977.

[54] Плато (Plato J., von). Creating Modern Probability. Its Mathematics, Physics and Philosophy in Historical Perspective. — Cambridge: Cambridge Univ. Press, 1994.

[55] Пойя (Pólya G.). Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem // Mathematische Zeitschrift. — 1920. — V. 8. — P. 171 — 181.

[56] Пуанкаре (Poincaré H.). Calcul des probabilités. — Paris: G. Carré, 1896.

[57] Пуанкаре (Poincaré H.). Sur le problème des trois corps et les équations de la dynamique. I, II // Acta Mathematica. — 1890. — V. 13. — P. 1 — 270.

[58] Сельванатан и др. (Selvanathan A., Selvanathan S., Keller G., Warrack B., Bartel H.). Australian Business Statistics. — Melbourne: Nelson, An International Thomson Publ. Co., 1994.

[59] Смолуховский (Smoluchowski M. R., von). Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen // Annalen der Physik. — 1906. — V. 21. — P. 756 — 780.

[60] Сэвидж (Savage L. J.). The Foundations of Statistics. — New York: Wiley; London :Chapman & Hall, 1954.

[61] Тодхантер (Todhunter I.). A History of the Mathematical Theory of Probability from the Time of Pascal to That of Laplace. — New York: Chelsea, 1949; — Первое издание: Cambridge: Macmillan, 1865.

[62] Уиман (Wiman A.). Über eine Wahrscheinlichkeitsaufgabe bei Kettenbruchentwicklungen // Akad. Förh. Stockholm. — 1900. — V. 57. — P. 829 — 841.

[63] Успенский, Семёнов (Uspensky V. A., Semenov A. L.). What are the gains of the theory of algorithms: basic developments connected with the concept of algorithm and with its application in mathematics // Algorithms in Modern Mathematics and Computer Science (Urgench, 1979). — Berlin etc.: Springer-Verlag, 1981. — P. 100 — 234. — (Lecture Notes in Computer Science; V. 122.)

[64] Файн (Fine T. L.). Theories of Probability. An Examination of Foundations. — New York — London: Academic Press, 1973.

[65] де Финетти (Finetti B., de). Sulle probabilità numerabili e geometriche // Istituto Lombardo. Accademia di Scienze e Lettere. Rendiconti (2). — 1928. — V. 61. — P. 817 — 824.

[66] де Финетти (Finetti B., de). Sulle funzioni a incremento aleatorio // Accademia Nazionale dei Lincei. Rendiconti (6). — 1929. — V. 10. — P. 163 — 168.

[67] де Финетти (Finetti B., de). Integrazione delle funzioni a incremento aleatorio // Accademia Nazionale dei Lincei. Rendiconti (6). — 1929. — V. 10. — P. 548 — 553.

[68] де Финетти (Finetti B., de). Probabilismo: saggio critico sulla teoria delle probabilità e sul valore della scienza. — Napoli: Perrella, 1931; // Logos. — 1931. — V. 14. — P. 163 — 219. — English transl.: // Erkenntnis. The International Journal of Analytic Philosophy. — 1989. — V. 31. — P. 169 — 223.

[69] де Финетти (Finetti B., de). Probability, Induction and Statistics. The Art of Guessing. — New York etc.: Wiley, 1972.

[70] де Финетти (Finetti B., de). Teoria delle probabilità: sintesi introduttiva con appendice critica. — V. 1, 2. — Turin: Einaudi, 1970. — English transl.: Theory of Probability: A Critical Introductory Treatment. — V. 1, 2. — New York etc.: Wiley, 1974, 1975.

[71] Фреше (Fréchet M.). Sur l'intégrale d'une fonctionnelle étendue à un ensemble abstrait // Bulletin de la Société Mathématique de France. — 1915. — V. 43. — P. 248 — 265.

[72] Хокинс (Hawkins T.). Lebesgue's Theory of Integration. Its Origin and Development. — Madison, Wis. — London: Univ. Wisconsin Press, 1970.

[73] Чёрч (Church A.) On the concept of a random sequence // American Mathematical Society. Bulletin. — 1940. — V. 46, № 2. — P. 130 — 135.

[74] Ширяев А. Н. Основы стохастической финансовой математики: В 2-хт. — М.: ФАЗИС, 1998.

[75] Эйнштейн (Einstein A.). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen // Annalen der Physik. — 1905. — V. 17. — P. 549 — 560.



图书文献资料

(第四章 ~ 第八章)

第四章

§1. A. H. 柯尔莫戈洛夫 "0 – 1" 律在他的书 [32] 中. 关于 E. 休伊特 (E. Hewitt) 和塞维奇 "0 – 1" 律, 亦见 A. A. 博罗夫科夫 (A. A. Боровков) [7], L. 布赖曼 (Л. Брейман) [8], R. B. 阿什 (R. B. Ash) [81].

§2~§4. 这里是 A. H. 柯尔莫戈洛夫和 A. Я. 辛钦得到的基本结果 (见 [32] 以及其中的文献). 亦见 B. B. 彼得罗夫 (W. W. Petrov) [53] 和, W. F. 斯托特 (W. F. Stout) [66]. 关于数论中的概率方法, 见 Й. 库比柳斯 (Й. Кубилюс)[①][36].

我们现在对于伯努利概形, 回忆 "强大数定律和重对数定律" 的历史.

关于强大数定律最早的工作, 出现在 E. 博雷尔如下论文中: "关于集合 [0, 1) 中数的正态性" (É. Borel. Les probabilittés dénombrables et leurs applications arithmétiqes // Rendiconti del Circolo Matematico di Palermo. — 1909. — V.27 — P.247~271). 如果使用 §3 例 2 的记号, 则对于量

$$S_n = \sum_{k=1}^{n} \left(I(\xi_k = 1) - \frac{1}{2} \right),$$

E. 博雷尔所得结果是: (关于勒贝格测度) 对于几乎一切 $\omega \in [0, 1)$, 存在 $N = N(\omega)$, 当 $n \geqslant N(\omega)$ 时, 有

$$\left| \frac{S_n(\omega)}{n} \right| \leqslant \frac{\ln(n/2)}{\sqrt{2n}}.$$

① 立陶宛数学家 (1921 年 ~), 立陶宛科学院院士, 立陶宛维尔纽斯大学校长. —— 译者

这样, 特别, 几乎必然 (处处) $S_n = o(n)$.

下一步由 F. 豪斯多夫 (F. Hausdorff, Grundzüge der Mengenlehre, — Leipzig: Veit, 1914), 证明了, 对于任意 $\varepsilon > 0$, 几乎必然 $S_n = o(n^{1/2+\varepsilon})$.

在 1914 年, G. H. 哈代 (G. H. Hardy) 和 J. E. 李特尔伍德 (J. E. Littlewood, Some Problems of Diophantine approximation//Acta Mathematica, — 1914. —, V.37, — P.155~239) 证明, 几乎必然 $S_n = O((n \ln n)^{1/2})$.

在 1922 年, H. D. 斯坦因豪斯 (H. D. Steinhaus, Les probabilités dénombrables et leur rappot à la théorie de la mesure//Fundamenta Mathematicae, — 1923. — V4, — P.286~310) 将 G. H. 哈代和 J. E. 李特尔伍德的结果精确化, 证明几乎必然, 有

$$\limsup_n \frac{S_n}{\sqrt{2n \ln n}} \leqslant 1.$$

1923 年 A. Я. 辛钦 (A. J. Khinchin, Über dyadische Brüche//Mathematische Zeitschrift — 1923. — V18, P.109~116) 断定, 几乎必然 $S_n = O(\sqrt{n \ln \ln n})$.

最后, 1924 年 A. Я. 辛钦 (A. J. Khinchin, Über einen Satzder Wahrscheinlichkeitsrechnung//Fundamenta Mathematicae, — 1924. — V.6. — P.9~20 到最终的结果 ("重对数定律"): 几乎必然, 有

$$\limsup_n \frac{S_n}{\sqrt{(n/2) \ln \ln n}} = 1.$$

(需要指出, 对于所考虑的情形,

$$\sigma^2 = \mathbf{E}\left[I(\xi_k = 1) - \frac{1}{2}\right]^2 = \frac{1}{4},$$

因为出现了因子 $n/2$, 而不是因子 $2n$; 对照 §4 定理 1 的提法).

如同 §4 提到的, 下一步就是对于广泛的独立随机变量类, 证明重对数定律, 这是由 A. H. 柯尔莫戈洛夫于 1922 年实现的 (A. N. Kolmogoroff, Über das Gesets des iterierten Logarithmus//Mathematische Annalen, — 1929. — V.101, — P.126~135).

§5. 关于这些问题, 见 B. B. 彼得罗夫 [92], A. A. 博罗夫科夫 [7], D. 达昆纳-卡斯特里 (D. Ducunna-Castelle) 和 M. 杜弗劳 (M. Duflo) [86].

第五章

§1~§3. 在叙述 (强) 平稳随机序列时, 借鉴了书籍: L. 布赖曼 [8], Я. Г. 希奈 (Я. Г. Синаи) [63], J. 兰珀蒂 (J. Lamperti) [38]. A. M. 加尔西亚 (A. M. Garsia) [12] 给出了最大遍历性定理的简单证明.

第六章

§1. 关于 (广义) 平稳随机序列理论见下列图书: Ю. А. 罗扎诺夫 (Ю. А. Розанов) [60], И. И. 基赫曼 (И. И. Гихман) 和 А. В. 斯科罗霍德 (А. В. Скороход) [13], [14]. А. Н. 柯尔莫戈洛夫常在讲课时举例 6.

§2. 关于正交随机测度和随机积分, 亦见 J. L. 杜布 (J. L. Doob) [20], И. И. 基赫曼和 А. В. 斯科罗霍德 [14], Ю. А. 罗扎诺夫 [60], R. B. 阿什和 M. F. 加德纳 M. F. Cardner [82].

§3. 谱表现 (2) 是 H. 克拉默 (H. Cramer) 和 M. 洛埃甫(M. Loéve) 得到的 (例如, 见 [42]). 在 А. Н. 柯尔莫戈洛夫的著作 [29] 中, (在其他术语下) 含这样的表现. 亦见如下著作: J. L. 杜布 [20], Ю. А. 罗扎诺夫 [60], R. B. 阿什和 M. F. 加德纳 M. F. Cardner [82].

§4. 在 E. J. 汉南 (E. J. Hannan) 的书 [71] 和 [72] 中, 有协方差函数和谱密度的统计估计问题的详细叙述.

§5~§6. 亦见下列图书: Ю. А. 罗扎诺夫 [60], J. 兰珀蒂 [38], И. И. 基赫曼和 А. В. 斯科罗霍德 [13], [14].

§7. 这部分内容, 是按 Р. Ш. 利普彩尔 (Р. Ш. Липцер) 和 А. Н. 施利亚耶夫的书 [41] 叙述的.

第七章

§1. 鞅论的多数基本结果是 J. L. 杜布得到的 [20]. 定理 1 包含在 P. -A. 麦耶 (P. -A. Meyer) 的专著 [47] 中. 亦见, P. -A. 麦耶的书 [48], Р. Ш. 利普彩尔和 А. Н. 施利亚耶夫的书 [41], И. И. 基赫曼和 А. В. 斯科罗霍德 [14], J. 扎克德 (J. Jacod) 和 А. Н. 施利亚耶夫[87].

§2. 定理 1 常称为 "关于自由选择的变换" 定理, [20]. 关于恒等式 (13) 和 (14), 以及 A. 瓦尔德基本恒等式, 见书 [9].

§3. 第一个不等式 (25), 是 А. Я. 辛钦于 1923 年在其论文 (A. J. Khintchine, Über dyadische Brüche//Mathematische Zeischrift — 1923. — V18, P.109~116) 中, 在证明 "重对数定律" 时得到的. 为说明是什么使 А. Я. 辛钦必然会得到该不等式, 我们提醒注意, E. 博雷尔和 F. 豪斯多夫证明强大数定律的概形 (亦见上面对第四章 §2~§4 的评述).

设 $\xi_1, \xi_2, \cdots$ 是独立同分布随机变量列, 且 $\mathbf{P}\{\xi_1 = 1\} = \mathbf{P}\{\xi_2 = -1\} = 1/2$ (伯努利概形), $S_n = \xi_1 + \cdots + \xi_n$.

E. 博雷尔的证明 "几乎必然 $S_n = o(n)$" 的实质如下: 由于对于任意 $\delta > 0$, 有

$$\mathbf{P}\left\{\left|\frac{S_n}{n}\right| \geqslant \delta\right\} \leqslant \frac{\mathbf{E}S_n^4}{n^4\delta^4} \leqslant \frac{3n^2}{n^4\delta^4} = \frac{3}{n^2\delta^4},$$

则当 $n \to \infty$ 时

$$\mathbf{P}\left\{\sup_{k\geqslant n}\left|\frac{S_k}{k}\right|\geqslant\delta\right\}\leqslant\sum_{k\geqslant n}\mathbf{P}\left\{\left|\frac{S_k}{k}\right|\geqslant\delta\right\}\leqslant\frac{3}{\delta^4}\sum_{k\geqslant n}\frac{1}{k^2}\to 0,$$

因此, 根据博雷尔 - 坎泰利引理 (第二章 §10), 几乎必然 $S_n/n \to 0$.

F. 豪斯多夫的证明, 对于任意 $\varepsilon > 0$, "几乎必然 $S_n = o(n^{1/2+\varepsilon})$", 是类似地进行的: 因为 $\mathbf{E}S_n^{2r} = O(n^r)$, 对于任意整数$r > 1/(2\varepsilon)$, 则当 $n \to \infty$ 时

$$\mathbf{P}\left\{\sup_{k\geqslant n}\left|\frac{S_k}{k^{1/2+\varepsilon}}\right|\geqslant\delta\right\}\leqslant\sum_{k\geqslant n}\mathbf{P}\left\{\left|\frac{S_k}{k^{1/2+\varepsilon}}\right|\geqslant\delta\right\}$$
$$\leqslant\frac{1}{\delta^{2r}}\sum_{k\geqslant n}\mathbf{E}\left|\frac{S_k}{k^{1/2+\varepsilon}}\right|^{2r}\leqslant\frac{c}{\delta^{2r}}\sum_{k\geqslant n}\frac{k^r}{k^{r+2\varepsilon r}}\to 0,$$

其中 c 是常数. 由此 (仍然根据博雷尔 - 坎泰利引理) 可得, 几乎必然

$$\frac{S_n}{n^{1/2+\varepsilon}}\to 0.$$

由以上的讨论可见, 证明的关键是对于概率 $\mathbf{P}\{|S_n| \geqslant t(n)\}$ 得到 "好" 估计, 其中对伯雷尔, $t(n) = n$, 豪斯多夫, $t(n) = n^{1/2+\varepsilon}$, (对哈代和李特尔伍德, $t(n) = (n\ln n)^{1/2}$).

正是为得到概率 $\mathbf{P}\{|S_n| \geqslant t(n)\}$ 的 "好" 估计, A. Я. 辛钦用到他的 "辛钦不等式" (25) (确切地说, 这些不等式中的第一个).

关于辛钦 (右和左) 不等式对于任意 $p > 0$ 的推导, 以及在 (25) 式中关于常数 A_p 和 B_p 之最优性的证明, 见 Г. 佩舍基尔 (Г. Пешкир) 和 A. H. 施利亚耶夫的综述性文章: Неравен-ства Хинчина и мартингальное расширение сферы их действия (辛钦不等式及其应用范围的扩展)//Успехи математических наук. —. 1995 年, Т.50, 5. стр. 3~62.

由 (25) 式的第一个不等式, 当 $p = 2m$ 时, A. Я. 辛钦得到, 对于任意 $t > 0$,

$$\mathbf{P}\{|X_n| > t\}\leqslant t^{-2m}\mathbf{E}|X_n|^{2m}\leqslant\frac{(2m)!}{2^m m!}t^{-2m}[X]_n^{2m}.$$

由斯特林公式

$$\frac{(2m)!}{2^m m!}\leqslant D\left(\frac{2}{\mathrm{e}}\right)^m m^m,$$

其中 $D = \sqrt{2}$. 因此, 若设 $m = [t^2/(2[X]_n^2)]$, 则有

$$\mathbf{P}\{|X_n| > t\}\leqslant D\left(\frac{2m[X]_n^2}{\mathrm{e}t^2}\right)^m\leqslant D\mathrm{e}^{-m}\leqslant D\exp\left\{1-\frac{t^2}{2[X]_n^2}\right\}$$
$$= D\mathrm{e}\exp\left\{\frac{-t^2}{2[X]_n^2}\right\} = c\exp\left\{\frac{-t^2}{2[X]_n^2}\right\},$$

其中 $c = De = \sqrt{2}e$.

由此估计, 可得不等式

$$\mathbf{P}\{|S_n| > t\} \leqslant e^{-\frac{t^2}{2c^2}}.$$

A. Я. 辛钦曾经利用该式, 证明 $S_n = O(\sqrt{n \ln\ln n})$ (a.c.).

在 Y. S. 乔 (Y. S. Chow) 和泰切尔 (H. Teicher) 的书中 [73], 有大量在这一节引用的不等式. 定理 2 属于 E. Lenglavt [39].

§4. 见 J. L. 杜布的专著 [20].

§5. 这里叙述的内容遵循如下读者的文章: Ю. М. 卡巴诺夫 (Ю. М. Кабанов), Р. Ш. 利普彩尔, А. Н. 施利亚耶夫 [26], H. -J. 恩格尔伯特 (H. -J. Engelbert), А. Н. 施利亚耶夫 [79], 以及奈维尤 (J. Neveu) 的书 [50]. 定理 4 和例是 Р. Ш. 利普彩尔提供的.

§6. 这里对 "绝对连续性和奇异性" 问题提供的处理方法, 以及所介绍的结果, 包含在 Ю. М. 卡巴诺夫, Р. Ш. 利普彩尔, А. Н. 施利亚耶夫的工作 [26] 中.

§7. 定理 1 和定理 2 世界语属于 А. А. 诺维科夫 (А. А. Новиков) [52]. 引理 1 是著名的吉尔萨诺夫 (Гирсанов) 定理的 "离散" 类似 (见 [41]).

§8. 亦见 Р. Ш. 利普彩尔和 А. Н. 施利亚耶夫的书 [91], J. 扎克德和 А. Н. 施利亚耶夫 [87], 在该书中讲述相当一般情形的随机过程论 (鞅, 半鞅 ······).

§9. 这里叙述遵从 [98],[100]. 对伊藤清 (Itô Kiyosi) 公式的推广, 所叙述的方法的发展, 见 H. 福尔默 (H. Föllmer), Ph. 普罗泰尔 (Ph. Protter) 和 А. Н. 施利亚耶夫的文章 [101].

§10. 关于保险中的鞅方法, 见格贝尔 (H. Gerber) 的书 [123]. 所作证明接近 [98] 中相应的证明.

§11~§12. 涉及鞅方法, 在金融数学和金融工程学中的应用问题, 较为详细的叙述, 见 [100].

§13. 最优停止规则主要专著是: Е. Б. 邓肯 (E. B. Dynkin) 和 А. А. 尤什克维奇 (А. А. Юшкевич) [102], H. R. 罗宾斯 (H. R. Robbins), D. 西格蒙德 (D. Sigmund) 和 Y. S. 乔 [59], А. Н. 施利亚耶夫 [78].

第八章

§1~§2. 关于马尔可夫链的定义和基本性质, 亦见下列著作: Е. Б. 邓肯和 А. А. 尤什克维奇 [102], Е. Б. 邓肯 [21], А. Д. 温策尔 (А. Д. Вентцель) [11], J. L. 杜布 [20], И. И. 基赫曼和 А. В. 斯科罗霍德 [14], L. 布赖曼 [8], 钟开莱 (Kai Lai Chung) [75],[120], D. 雷夫尤兹 (D. Revuz) [117].

§3~§7. 关于马尔可夫链的, 极限, 遍历和平稳概率分布问题, 见 А. Н. 柯尔莫戈洛夫文章 [28], 和如下书籍: W. 费勒 (W. Feller) [69], А. А. 博罗夫科夫 [7], [104], R.

B. 埃什 (R. B. Ashi), 钟开莱 [120], D. 雷夫尤兹 [117], E. B. 邓肯和 A. A. 尤什克维奇 [102].

§8. 简单随机游动, 是最简单的马尔可夫链的经典例子, 对此情形曾经发现了许多规律性 (例如, 常返性于非常返性, 遍历性等等). 有关内容在许多图书中都有, 例如, 上面引用过的书 [7],[80],[120],[117].

§9. 统计序列分析问题 (A. 瓦尔德 [9], M. de 格鲁特 (M. de Gloot) [18], S. 萨克斯 (S. Sacks) [22], A. H. 施利亚耶夫 [78]), 决定了对于最优停时问题的兴趣. 关于马尔可夫链的最优停时规则本身, 有如下著作: E. B. 邓肯和 A. A. 尤什克维奇 [102], A. H. 施利亚耶夫 [78]), P. 比林斯利 (P. Billingsley) [106] 的某些章节. 最优停时问题的鞅方法, 见 H. R. 罗宾斯, D. 西格蒙德和 Y. S. 乔的专著 [59].

概率的数学理论形成的简史

该简史是本书的作者, 为 A. H. 柯尔莫戈洛夫的专著《概率论的基本概念》[32] 第三版, 作为补充而写的.



参考文献

[1] Александров П. С. Введение в общую теорию множеств и функций. — М.: Гостехиздат, 1948.

[2] Александрова Н. В. Математические термины. — М.: Высшая школа, 1978.

[3] Бернштейн С. Н. О работах П. Л. Чебышева по теории вероятностей//Научное наследие П. Л. Чебышева. Вып.1: Математика. — 1945. — С. 59 — 60.

[4] Бернштейн С. Н. Теория вероятностей. — 4-е изд. — М.: Гостехиздат, 1946.

[5] Биллингсли П. Сходимость вероятностных мер. — М.: Наука, 1977.

[6] Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. — 3-е изд. — М.: Наука, 1983.

[7] Боровков А. А. Теория вероятностей. — 3-е изд. — М.: УРСС, 1999.

[8] Брейман (Breiman L.). Probability. — Reading, MA: Addison-Wesley, 1968.

[9] Вальд А. Последовательный анализ. — М.: Физматгиз, 1960.

[10] Ван дер Варден Б. Л. Математическая статистика. — М.: ИЛ, 1960.

Под номерами 85   101 идет литература, добавленная во втором издании к той, которая была приведена в первом издании книги. Под номерами 102 — 136 идет литература, добавленная в настоящем издании.

[11] Вентцель А. Д. Курс теории случайных процессов. — М.: Наука, 1975.

[12] Гарсиа (Garsia A. M.). A simple proof of E. Hopf's maximal ergodic theorem // Journal of Mathematics and Mechanics. — 1965. — V.14, № 3. — P. 381 – 382.

[13] Гихман И. И., Скороход А. В. Введение в теорию случайных процессов. — М.: Наука, 1977.

[14] Гихман И. И., Скороход А. В. Теория случайных процессов: В3т. — М.: Наука, 1971 — 1975. 汉译本:《随机过程论》, 第一卷 (邓永录等译), 第二卷 (周概容译), 1986.

[15] Гнеденко Б. В. Курс теории вероятностей. — 6-е изд. — М.: Наука, 1988. 汉译本:《概率论教程》(丁寿田), (第三版), 高等教育出版社, 1961.

[16] Гнеденко Б. В., Колмогоров А. Н. Предельные распределения для сумм независимых случайных величин. — М.: Л.: Гостехиздат, 1949. 汉译本:《相互独立随机变量之和的极限分布》(王寿仁译), 科学出版社, 1950.

[17] Гнеденко Б. В., Хинчин А. Я. Элементарное введение в теорию вероятностей. — 9-е изд. — М.: Наука, 1982.

[18] Де Гроот М. Оптимальные статистические решения. — М.: Мир, 1974.

[19] Дохерти (Doherty M.). An amusing proof in fluctuation theory// Combinatorial Mathematics, III: Proceedings of the Third Australian Conference, Univ. Queensland, St. Lucia 1974. — Berlin etc.: Springer-Verlag, 1975. — P.101 — 104. — (Lecture Notes in Mathematics; V.452.)

[20] Дуб Дж. Л. Вероятностные процессы. — М.: ИЛ, 1956.

[21] Дынкин Е. Б. Марковские процессы. — М.: Физматгиз, 1963.

[22] Закс Ш. Теория статистических выводов. — М.: Мир, 1975.

[23] Ибрагимов И. А., Линник Ю. В. Независимые и стационарно связанные величины. — М.: Наука, 1965.

[24] Ибрагимов И. А., Розанов Ю. А. Гауссовские случайные процессы. — М.: Наука, 1970.

[25] Исихара А. Статистическая физика. — М.: Мир, 1973.

[26] Кабанов Ю. М., Липцер Р. Ш., Ширяев А. Н. К вопросу об абсолютной непрерывности и сингулярности вероятностных мер// Математический сборник. — 1977, — Т. 104. № 2. — С. 227 — 247.

[27] Кемени Дж., Снелл Дж. Конечные цепи Маркова. — М.: Наука, 1970.

[28] Колмогоров А. Н. Цепи Маркова со счетным числом возможных состояний//Бюллетень МГУ. — 1937. — Т.1, № 3. — С. 1 — 16.

[29] Колмогоров А. Н. Стационарные последовательности в гильбертовском пространстве//Бюллетень МГУ. — 1941. — Т. 2, № 6. — С. 1 — 40.

[30] Колмогоров А. Н. Роль русской науки в развитии теории вероятностей//Ученые записки МГУ. — 1947. — Вып. 91. — С. 53 — 64. 汉译本:《概率论》(见 “数学, 它的内容、方法和意义” (卷2)), 科学出版社, 1961.

[31] Колмогоров А. Н. Теория вероятностей//Математика, ее содержание, методы и значение. — М.: Изд-во АН СССР, 1956. — Т. II. — С. 252 — 284.

[32] Колмогоров А. Н. Основные понятия теории вероятностей. — М.; Л.: ОНТИ, 1936; 2-е изд. М.: Наука, 1974; 3-е изд. М.: Фазис, 1998. 汉译本:《概率论的基本概念》(丁寿田译), 商务印书馆, 1952.

[33] Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа. — 6-е изд. —М.: Наука, 1989.

[34] Колчин В. Ф., Севастьянов Б. А., Чистяков В. П. Случайные размещения. — М.: Наука, 1976.

[35] Крамер Г. Математические методы статистики. — 2-е изд. — М.: Мир, 1976. 汉译本:《统计学数学方法》(魏宗舒译), 上海科技出版社, 1983.

[36] Кубилюс Й. Вероятностные методы в теории чисел. — Вильнюс: Гос. изд-во полит. и науч. лит. ЛитССР, 1959.

[37] Ламперти Дж. Вероятность. — М.: Наука, 1973.

[38] Ламперти (Lamperti J.). Stochastic Processes. — Aarhus Univ., 1974. — (Lecture Notes Series; № 38).

[39] Ленгляр (Lenglart E.). Relation de domination entre deux processus// Annales de l'Institut H. Poincaré Sect. B. (N. S.). — 1977. — V.13, № 2. — P. 171 — 179.

[40] Леонов В. П., Ширяев А. Н. К технике вычисления семиинвариантов// Теория вероятностей и ее применения. — 1959. — Т. IV, вып. 2. — С. 342 — 355.

[41] Липцер Р. Ш., Ширяев А. Н. Статистика случайных процессов. — М.: Наука, 1974. 汉译本:《随机过程统计》.

[42] Лоэв М. Теория вероятностей. — М.: ИЛ, 1962. 汉译本:《概率论》, 上册, (梁文骐译), 科学出版社, 1966.

[43] Марков А. А. Исчисление вероятностей. — 3-е изд. — СПб., 1913.

[44] Майстров Д. Е. Теория вероятностей (исторический очерк). — М.: Наука, 1967.

[45] Математика XIX века/ Под ред. А. Н. Колмогорова и А. П. Юшкевича. — М.: Наука, 1978.

[46] Мешалкин Л. Д. Сборник задач по теории вероятностей. — М.: Изд-во МГУ, 1963. 汉译本:《概率论习题集》(盛骤等译), 高等教育出版社, 1984.

[47] Мейер (Meyer P.-A.). Martingales and Stochastic Integrals. I. — Berlin etc.: Springer-Verlag, 1972. — (Lecture Notes in Mathematics; V. 284).

[48] Мейер П. -А. Вероятность и потенциалы. — М.: Мир, 1973.

[49] Неве Ж. Математические основы теории вероятностей. — М.: Мир, 1969.

[50] Невё (Neveu J.). Discrete-Parameter Martingales. — Amsterdam etc.: North-Holland, 1975.

[51] Нейман Ю. Вводный курс теории вероятностей и математической статистики. — М.: Наука, 1968.

[52] Новиков А. А. Об оценках и асимптотическом поведении вероятностей непересечения подвижных границ суммами независимых случайных величин//Известия АН СССР. Серия математическая. — 1980. — Т. 40, вып. 4. — С. 868 — 885.

[53] Петров В. В. Суммы независимых случайных величин. — М.: Наука, 1972

[54] Прохоров Ю. В. Асимптотическое поведение биномиального распределения// Успехи математических наук. — 1953. — Т. VIII, вып. 3(55). — С. 135—142.

[55] Прохоров Ю. В. Сходимость случайных процессов и предельные теоремы теории вероятностей// Теория вероятностей и ее применения. — 1956. — Т. I, вып. 2. — С. 177—238.

[56] Прохоров Ю. В., Розанов Ю. А. Теория вероятностей. — 2-е изд. — М.: Наука, 1973.

[57] Рамачандран Б. Теория характеристических функций. — М.: Наука, 1975.

[58] Реньи (Rényi A.) Probability Theory. — Amsterdam: North-Holland, 1970.

[59] Роббинс Г., Сигмунд Д., Чао И. Теория оптимальных правил остановки. — М.: Наука, 1977.

[60] Розанов Ю. А. Стационарные случайные процессы. — М.: Физматгиз, 1963.

[61] Сарымсаков Т. А. Основы теории процессов Маркова. — М.: Гостехиздат, 1954.

[62] Севастьянов Б. А. Ветвящиеся процессы. — М.: Наука, 1971.

[63] Синай Я. Г. Введение в эргодическую теорию. — Ереван: Изд-во Ереван. ун-та, 1973.

[64] Сираждинов С. Х. Предельные теоремы для однородных цепей Маркова. — Ташкент: Изд-во АН УзССР, 1955.

[65] Справочник по теории вероятностей и математической статистике/ Под ред. В. С. Королюка. — Киев: Наукова думка, 1978.

[66] Стоут (Stout W. F.). Almost Sure Convergence. — New York etc.: Academic Press, 1974.

[67] Теорія имовірностей. — Киів: Вища школа, 1976.

[68] Тодхантер (Todhunter I.). A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace. — London: Macmillan, 1865.

[69] Феллер В. Введение в теорию вероятностей и ее приложения: В 2-х т. — М.: Мир, 1984. 汉译本:《概率论导引》(胡迪鹤译), 人民邮电出版社, 2006.

[70] Халмош П. Теория меры. — М.: ИЛ, 1953. 汉译本:《测度论》(王建华译), 科学出版社.

[71] Хеннан Э. Анализ временных рядов. — М.: Наука, 1964.

[72] Хеннан Э. Многомерные временные ряды. — М.: Мир, 1974.

[73] Чао, Тейчер (Chow Y. S., Teicher H.). Probability Theory. Independence, Interchangeability, Martingales. — 3rd ed. — New York: Springer-Verlag, 1997.

[74] Чебышев П. Л. Теория вероятностей: Лекции акад. П. Л. Чебышева, читанные в 1879, 1880 гг./ Издано А. Н. Крыловым по записи А. М. Ляпунова. — М.; Л., 1936.

[75] Чжун Кай-лай. Однородные цепи Маркова. — М.: Мир, 1964.

[76] Ширяев А. Н. Случайные процессы. — М.: Изд-во МГУ, 1972.

[77] Ширяев А. Н. Вероятность, статистика, случайные процессы: В 2-х т. — М.: Изд-во МГУ, 1973 — 1974.

[78] Ширяев А. Н. Статистический последовательный анализ. — 2-е изд. — М.: Наука, 1976.

[79] Энгельберт, Ширяев (Engelbert H.-J., Shiryaev A. N.). On the sets of convergence of generalized submartingales // Stochastics. — 1979. — V. 2, № 3. — P. 155 — 166.

[80] Эш (Ash R. B.). Basic Probability Theory. — New York etc.: Wiley, 1970.

[81] Эш (Ash R. B.). Real Analysis and Probability.—New York etc.: Academic Press, 1972.

[82] Эш, Гарднер (Ash R. B., Gardner M. F.). Topics in Stochastic Processes. — New York etc.: Academic Press, 1975.

[83] Яглом А. М., Яглом И. М. Вероятность и информация. — 3-е изд. — М.: Наука, 1973. 汉译本:《概率与信息》, 科学出版社.

[84] Гринвуд, Ширяев (Greenwood P. E., Shiryayev A. N.). Contiguity and the Statistical Invariance Principle. — London: Gordon & Breach, 1985.

[85] Дадли (Dudley R. M.)Distances of probability measures and random variables// Annals of Mathematical Statistics. — 1968. — V.39, № 5. — P. 1563 — 1572.

[86] Дакуна-Кастелль, Дюфло (Dacunha-Castelle D., Duflo M.). Probabilités et statistiques: 1, 2. — Paris: Masson. — 1: Problèmes à temps fixe. — 1982; — 2: Problèmes à temps mobile. — 1983. — Перев. на англ. яз.: Probability and Statistics: V. I, II. — Berlin etc.: Springer-Verlag, 1986.

[87] Жакод Ж., Ширяев А. Н. Предельные теоремы для случайных процессов: В 2-х т. — М.: Физматлит, 1994.

[88] Золотарев В. М. Современная Теория суммирования независимых случайных величин. — М.: Наука, 1986.

[89] Ле Кам (Le Cam L.). Asymptotic Methods in Statistical Decision Theory. — Berlin etc.: Springer-Verlag, 1986.

[90] Лизе, Вайда (Liese F., Vajda I.). Convex Statistical Distances. — Leipzig: Teubner, 1987.

[91] Липцер Р. Ш., Ширяев А. Н. Теория мартингалов. — М.: Наука, 1986.

[92] Петров В. В. Предельные теоремы для сумм независимых случайных величин. — М.: Наука, 1987.

[93] Поллард (Pollard D.). Convergence of Stochastic Processes. — Berlin etc.: Springer-Verlag, 1984.

[94] Пресман Э. Л. О сближении по вариации распределения суммы независимых бернуллиевских величин с пуассоновским законом// Теория вероятностей и ее применения. — 1985. — Т. XXX, вып. 2. — С. 391 — 396.

[95] Розанов Ю. А. Теория вероятностей, случайные процессы и математическая статистика. — М.: Наука, 1985.

[96] Ротарь В. И. Кобобщению теоремы Линдеберга — Феллера// Математические заметки.— 1975. — Т.18, вып. 1. — С. 129 — 135.

[97] Севастьянов Б. А. Курс теории вероятностей и математической статистики. — М.: Наука, 1982.

[98] Ширяев (Shiryayev A. N.) Probability. — 2nd ed. — Berlin etc.: Springer-Verlag, 1995.

[99] Ширяев (Shirjayev A. N.) Wahrscheinlichkeit. — Berlin: VEB Deutscher Verlag der Wissenschaften, 1988.

[100] Ширяев А. Н. Основы стохастической финансовой математики: В 2-х т. — М.: ФАЗИС, 1998.

[101] Фёллмер, Проттер, Ширяев (Föllmer H., Protter Ph., Shiryaev A. N.). Quadratic covariation and an extension of Itô's formula//Bernoulli. — 1995. — V.1, № 1/2. — P. 149 — 170.

[102] Дынкин Е. Б., Юшкевич А. А. Теоремы и задачи о процессах Маркова. — М.: Наука, 1967.

[103] Гнеденко, Колмогоров (Gnedenko B. V., Kolmogorov A. N). Limit Distributions for Sums of Independent Random Variables. — Reading, MA, etc.: Addison-Wesley, 1954.

[104] Боровков А. А. Эргодичность и устойчивость случайных процессов. — М.: УРСС, 1999.

[105] Гриммет, Стирзакер (Grimmet G. R., Stirzaker D. R.). Probability and Random Processes. — Oxford: Clarendon Press, 1993.

[106] Биллингсли (Billingsley P.). Probability and Measure. — 3rd ed. — New York: Wiley, 1995.

[107] Боровков А. А. Математическая статистика. — М.: Наука, 1984.

[108] Дарретт (Durrett R.). Probability: Theory and Examples. — Pacific Grove, CA: Wadsworth & Brooks/Cole, 1991.

[109] Дарретт (Durrett R.). Stochastic Calculus. — Boca Raton, FL: CRC Press, 1996.

[110] Дарретт (Durrett R.). Brownian Motion and Martingales in Analysis. — Belmont, CA: Wadsworth International Group, 1984.

[111] Калленберг (Kallenberg O.). Foundations of Modern Probability. — 2nd ed. — New York: Springer-Verlag, 2002.

[112] Карлин, Тейлор (Karlin S., Taylor H. M.). A First Course in Stochastic Processes. — 2nd ed. — New York etc.: Academic Press, 1975.

[113] Кашин Б. С., Саакян А. А. Ортогональные ряды. — 2-е изд. — М.: АФЦ, 1999.

[114] Жакод, Проттер (Jacod J., Protter Ph.). Probability Essentials. — Berlin etc.: Springer-Verlag, 2000.

[115] Нётс (Neuts V. F.) Probability. — Boston, MA: Allyn & Bacon, 1973.

[116] Плато (Plato J.). Creating Modern Probability. — Cambridge: Cambridge Univ. Press, 1998.

[117] Ревюз Д. Цепи Маркова. — М.: РФФИ, 1997.

[118] Уильямс (Williams D.). Probability with Martingales. — Cambridge: Cambridge Univ. Press, 1991.

[119] Холл, Хейде (Hall P., Heyde C. C.). Martingale Limit Theory and Its Applications. — New York etc.: Academic Press, 1980.

[120] Чжун Кай-Лай (Chung Kai Lai). Elementary Probability Theory with Stochastic Processes. — 3rd ed. — Berlin etc.: Springer-Verlag, 1979. 汉译本:《初等概率论和随机过程》(吕乃刚等译), 人民教育出版社.

[121] Математическая энциклопедия: В 5 т./Гл. ред. И. М. Виноградов. — М.: Советская энциклопедия, 1977 — 1985. 汉译本:《数学百科全书》, 第一卷 ~ 第五卷, 科学出版社, 1984 — 2000.

[122] Стиглер (Stigler S. M.).The History of Statistics: The Measurement of Uncertainty Before 1900. — Cambridge: Belknap Press of Harvard Univ. Press, 1986.

[123] Гербер Х. Математика страхования жизни. — М.: Мир, 1995.

[124] Эренфесты П. и Т. (Ehrenfest P., Ehrenfest T.). Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem//Physikalische Zeitschrift. — 1907. — V.8. — P. 311 — 314.

[125] Теория вероятностей и математическая статистика: энциклопедия/ Гл. ред. Ю. В. Прохоров. — М.: Большая Российская энциклопедия, 1999.

[126] Вольфрам (Wolfram S.). The Mathematica® Book. — 4th ed. — Champaign; Cambridge: Wolfram Media; Cambridge Univ. Press, 1999.

[127] Дуб (Doob J. L.). What is a martingale?//The American Mathematical Monthly. — 1971. — V.78. — P. 451 — 463.

[128] Синай Я. Г. Курс теории вероятностей. — М.: Изд-во МГУ, 1985. — 2-е изд., 1986.

[129] Синай (Sinaĭ Ya. G.). Topics in Ergodic Theory. — Princeton, NJ: Princeton Univ. Press, 1999. — (Princeton Mathematical Series; V. 44.)

[130] Вальтерс (Walters P.) An Introduction to Ergodic Theory. — New York etc.: Springer-Verlag, 1982.

[131] Булинский А. В., Ширяев А. Н. Теория случайных процессов. — М.: Физматлит, 2003.

[132] Хмаладзе Э. В. Мартингальный подход в теории непараметрических критериев согласия// Теория вероятностей и ее применения. — 1981. — Т. XXVI. вып. 2. — С. 246 — 265.

[133] Гамильтон (Hamilton J. B.). Time Series Analysis. — Princeton, NJ: Princeton Univ. Press, 1994.

[134] Бернулли Я. О законе больших чисел. — Ч. 4: Искусство предположений. — М.: Наука, 1986.

[135] Лукач Е. Характеристические функции. — М.: Наука, 1979.

[136] Хренников (Khrennikov A.). Interpretations of Probability. — Utrecht: VSP, 1999.



名词索引

(汉语拼音为序, 右上角带 “1” 者为第一卷页码)

A

B

G

K

L

M

N

P

T

W

人名表

(汉语拼音为序)

A

阿什	R. B. Ash	Р. Эш
埃尔米特	Ch. Hermite	Ш. Эрмит
埃伦弗斯特	P. Ehrenfest	П. Эренфест
埃伦弗斯特	T. Ehrenfest	Т. Эренфест
埃森	C. G. Esseen	К. Г. Эссеен
爱因斯坦	A. Einstein	А. Эйнштейн
埃什	R. B. Ashi	Р. Б. Эш
埃舍	Esher	Эшер
埃特麦迪	N. Etemady	Н. Этемади
奥塔维安尼	Ottawiani	Оттавиани

B

巴特利特	M. S. Bartlett	М. С. Бартлетт
巴拿赫	S. Banach	С. Банах
巴彻里耶	L. bachelier	Л. Башелье
邦弗尔罗尼	Bonferroni	Бонферрони
鲍斯	S. N. Bose	Ш. Бозе
贝尔	A. G. Bell	А. Г. Белл
贝尔	R. L. Baire	Р. Л. Бэр

贝尔曼	R. E. Bellman	Р. Э. Беллман
贝克尔雷尔	A. H. Becquerel	А. Х. Беккерелем
贝里	A. C. Berry	А. С. Берри
贝塞尔	F. W. Bessel	Ф. В. Бессель
贝叶斯	T. Bayes	Т. Байес
比林斯利	P. Billingsley	П. Биллингсли
彼得罗夫	W. W. Petrov	В. В. Петров
毕达哥拉斯	Pythagoras	Пифагор
波拉德	D. Pollard	Д. Пполлард
波利亚	G. Pólya	Д. Пойа
伯恩斯坦	S. N. Bernstein	С. Н. Бернштейн
伯克霍尔德	D. L. Burkholder	Д. Л. Буркхольдер
伯克霍夫	G. D. Birkhoff	Дж. Д. Биркгоф
伯努利	D. Benoulli	Д. Бернулли
伯努利	J. Benoulli	Я. Бернулли
泊松	S. D. Poisson	С. Д. Пуассон
博赫纳	S. Bochner	С. Бохнер
博雷尔 (波莱尔)	E. Borel	Э. Борель
博利舍夫	L. N. Bolishev	Л. Н. Большев
博罗夫科夫	A. A. Borowkov	А. А. Боровков
布罗登	T. Brodén	Т. Броден
布尔	G. Boole	Дж. Буль
布耳曼	G. BoHlmann	Г. Больман
布耳兹曼	L. Boltzmann	Л. Больцман
布莱克韦尔	D. H. Blackwell	Д. Блэкуэлл
布赖曼	L. Breiman	Л. Брейман
布朗	E. T. Brown	Э. Т. Броун
布雷	J. R. Bray	Я. Р. Брэй
布罗吉	U. Broggi	У. Брогги
布洛赫	A. Bloch	А. Блох
布尼亚科夫斯基	A. J. Buniakowsky	А. Я. Буняковский
布西	R. S. Busey	Р. С. Вьюсу

C~D

查普曼	D. G. Chapman	Д. Г. Чепман
达布	J. G. Darboux	Ж. Г. Дарбу
达德利	R. M. Dudley	Р. Дадли

达昆纳－卡斯特里	D. Ducunna-Castelle	Д. Дакуна-Кастелле
达雷特	R. Durrett	Р. Даррет
戴维斯	H. T. Davis	Х. Т. Дэвис
德沃列茨基	—	Дворецкий
德格鲁特	M. H. deGroot	М. Де Гроот
邓肯	E. B. Dynkin	И. Б. Дынкин
狄拉克	P. A. M. Dirac	П. А. М. Дирак
狄利克雷	P. G. L. Dirichlet	П. Г. А. Дирихле
笛卡儿	R. Descartes	Р. Декарт
棣莫弗	A. Dé Moivre	А. Дé Муавр
杜布	J. L. Doob	Дж. Л. Дуб
杜弗劳	M. Duflo	М. Дюфло

E

恩格尔伯特	H.-J. Engelbert	Г.-Ю. Энгельберт

F

法图	F. Fatou	П. фату
范德瓦尔登	B. L. van der Waerden	Б. Л. Ван дер Варден
范基	Fan Ky	Ки Фан
费勒	W. Feller	В. Феллер
费马	P. Fermat	П. Ферма
费内提	B. de Finetti	Б. де Финетти
费希尔 (费歇尔)	R. A. Fisher	Р. А. Фишер
费叶尔	L. Féjer	Л. Фейер
弗雷歇	M. Fréchet	Ф. Фреше
弗米	E. Fermi	Э. Ферми
冯・诺伊曼	J. von Neuman	Дж. Фон Нейман
福尔默	H. Föllmer	Г. Фёллмер
福明	S. W. Fomin	С. В. Фомин
傅比尼	G. Fubini	Г. фубини
费里德曼	H. M. Friedman	Г. М. Фельдман
傅里叶	J. B. J. Fourier	Ж. Б. Ж. Фурье

G

盖尔登	J. A. H. Gyldén	Дж. А. Х. Гюлден
冈贝尔	E. J. Gumbel	Э. Гумбель
高斯	G. F. Gauss	К. Ф. Гаусс
格贝尔	H. Gerber	Х. Гербер
格拉姆	G. P. Gram	Г. П. Грам
格里米特	G. R. Grimmit	Дж. Гриммет
格里汶科	W. I. Glivenko	В. И. Гливенко
格林伍得	P. E. Greenword	П. Е. Гринвуд
格鲁特	M. de Gloot	М. де Гроот
格涅坚科	B. V. Gnedenko	Б. В. Гнеденко
哥塞特	W. S. Gosset	В. С. Госсет

H

哈恩	H. Hahn	Г. Хан
哈尔	A. Haar	А. Хаар
哈伊克	J. Hajek	Дж. Гаек
哈代	G. H. Hardy	Г. Х. Харди
哈尔默斯	P. R. Halmos	П. Халмош
哈密顿	W. R. Hamilton	У. Р. Гамильтон
哈特曼	P. Hartman	П. Хартман
汉南	E. J. Hannan	Э. Дж. Хеннан
海林格	E. Helinger	Э. Хеллингер
海涅	H. E. Heine	Г. Э. Нейне
豪斯多夫	F. Hausdorff	Ф. Хаусдрф
赫尔德	O. L. Hölder	О. Л. Гёльдер
赫尔格洛茨	G. Herglotz	Г. Герглотц
赫利	E. Helly	Э. Хелли
惠更斯	Ch. Huyghens	Х. Гюйгенс
霍普夫	H. Hopf	Х. Хопф
霍奇	W. W. D. Hodge	У. В. Д. Ходе

J

吉尔萨诺夫		И. Гирсанов
基赫曼	E. E. Gihman	И. И. Гихман
加德纳	M. F. Cardner	М. Гарднер
加尔西亚	A. M. Garsia	А. М. Гарсиа
角谷	S. Kakutani	С. Какутани

K

卡巴诺夫	Y. M. Kabanov	Ю. М. Кабанов
卡尔达诺	G. Cardanno	Дж. Кардано
卡尔莱曼	T. Carleman	Т. Карлеман
卡尔曼	R. E. Kalman	Р. И. Калман
卡拉泰奥多里	C. Carathéodory	К. Каратеодори
卡利卡尼尼	C. Calcagnini	Ч. Кальканини
卡姆	L. Le Cam	Л. Ле Кам
凯麦尼	J. G. Kemeny	Дж. Кемени
坎泰利	F. P. Cantelli	Ф. П. Кантелли
康托尔	G. Cantor	Г. Контор
考克斯	J. C. Cox	Дж. К. Кокс
柯尔莫戈洛夫	A. N. Kolmogorov	А. Н. Колмогоров
柯西	A. L. Cauchy	О. Л. Коши
科尔钦	W. F. Kerchen	В. Ф. Колчин
克里克伯格	Krickbeg	Крикберг
克拉默	H. Cramer	Г. Крамер
克罗内克	L. Kronecker	Л. Кронекер
库尔贝克	S. Kullback	С. Кульбак

L

拉奥	C. R. Rao	С. Р. Рао
拉德马赫	H. Rademacher	Г. А. Радемахер
拉东	J. Radon	Дж. Радон
拉格朗日	J. L. Lagrange	Ж. Л. Лагранж
拉曼钱德兰	B. Lamanchandran	Б. Раманчадран
拉普拉斯	P. S. Laplace	П. С. Лаплас
莱布尼茨	G. W. Lebniz	Г. В. Лейбниц
莱斯	F. Liese	Ф. Лизе
莱维	E. E. Levy	И. И. Леви
兰珀蒂	J. Lamperti	Дж. Ламперти
勒贝格	H. L. Lebesgue	А. Л. Лебег
雷夫尤兹	D. Revuz	Д. Ревюз
雷内伊	A. Rènyi	А. Ренуи
黎曼	G. F. B. Riemann	Г. Ф. Б. Риман
李雅普诺夫	A. M. Lyapunov	А. М. Ляпунов

李特尔伍德	J. E. Littlewood	Дж. И. Литтлвуд
里斯	F. Riesz	Ф. Рисс
利普彩尔	R. S. Lipchail	Р. Ш. Липцер
利普希茨	R. O. S. Lipschitz	Р. Липшиц
刘维尔	J. Liouville	Дж. Лиувилль
列昂诺夫	W. P. Leonov	В. П. Леонов
列维	P. P. Lévy	П. П. Леви
林德伯格	J. W. Linderberg	Дж. У. Линдеберг
林格拉特	E. Linglart	Э. Лингляр
洛埃甫	M. Loéve	М. Лоэв
洛必达	L' Hospital	Г. Лопиталь
伦德伯格	G. A. Lundeberg	Г. А. Лундберг
鲁宾斯坦	M. Rubinstein	М. Рубинштейн
罗塔里	W. E. Rotari	В. И. Ротарь
罗扎诺夫	Yu. A. Rozanov	Ю. А. Розанов
罗宾斯	H. R. Robbins	Г. Р. Робенс
罗斯	R. A. Ross	Р. А. Росс

M

马尔可夫	A. A. Markov	А. А. Марков
马歇尔	A. W. Marshall	А. В. Маршалл
马钦凯维奇	J. Matcinkiewicz	Й. Марцинкевич
马哈拉诺比斯	P. S. Mahalanobis	П. С. Махаланобис
迈斯特罗夫	D. I. Mastrov	Д. Е. Майстров
麦耶	P. -A. Meyer	П. -А. Мейер
麦克米兰	B. McMillan	Б. Макмиллан
麦克斯韦	J. C. Maxwell	Д. К. Максвелл
曼	H. B. Mann	Х. Б. Манн
梅沙尔金	L. D. Mesharlkin	Л. Д. Мешалкин
门罗	Monroe	Монро
米泽斯	R. Mises	Р. Миэесу
闵可夫斯基	H. Minkowski	Г. Минковский
默瑟	J. Mercer	Ж. Мерсер

N

奈曼	Yu. Neyman	Ю. Нейман
奈维尤	J. Neveu	Р. Невё
尼科迪姆	O. M. Nikodym	О. М. Никодим
牛顿	I. Newton	И. Ньютон
诺维科夫		А. А. Новиков

O

欧几里得	Euclid	Евклид
欧拉	L. Euler	Л. Эйлер

P

帕赞	E. Parzen	И. Парзен
帕利	W. Pauli	В. Паули
帕乔里	L. Papcioli	Л. Пачоли
帕塞瓦尔	M. A. Parseval	М. А. Пасеваль
帕斯卡	B. Pascal	Б. Паскаль
庞加莱	J. H. Poincaré	Ж. Ан. Пуанкаре
蒲丰 (布丰)	G. L. L. Buffon	Ж. Л. Л. Бюффон
普拉托	Jan von Plato	Ян Фон Плато
普拉特	Pratt	Пратт
普雷斯曼	I. L. Pressman	Э. Л. Пресман
普罗泰尔	Ph. Protter	Ф. Проттер
普罗霍罗夫	Yu. V. Prokhorov	Ю. В. Прохров

Q~R

奇斯佳科夫	W. P. Qisjiakov	В. П. Чистяков
乔	Y. S. Chow	Ю. Ш. Чао
切比雪夫 (切贝绍夫)	P. L. Chebyshev	П. Л. Чебышёв
齐格蒙特	A. Zygmund	А. Зигмунд
乔尔奇	A. Church	А. Чёрч
切萨罗	E. Cesàro	Э. Чеэаро
切尔诺夫		Чернов
茹尔边科		Журбенко

S

萨雷姆萨科夫	T. A. Saramsakov	Т. А. Сарымсаков
塞维治	I. R. Sevage	И. Р. Сэвидж
塞维治	L. J. Sevage	Л. И. Сэвидж
塞格	G. Szegö	Г. Сеге
沙利耶	C. L. Charlier	К. Л. Шарлье
绍德尔	A. Schauder	А. Шаудер
舍伊宁	O. W. Sherining	О. В. Шейнин
施利亚耶夫	A. N. Shiryaev	А. Н. Ширяев
施密特	E. Schimidt	Э. Шмидт
施瓦茨	L. Schwarz	Л. Шварц
斯莫卢霍夫斯基	M. Smoluchowski	М. Смолуховский
斯梯格列尔	S. M. Stigler	С. М. Стиглер
斯蒂尔切斯	T. J. Stieltjes	Т. И. Стилтьес
斯捷克洛夫	V. A. Steclov	В. А. Стеклов
斯科罗霍德	A. W. Skorokhod	А. В. Скороход
斯莱皮恩	P. Slepian	П. Слепян
斯鲁斯基	E. Slutsky	Е. Е. Слуцкий
斯米尔诺夫	N. V. Smirnov	Н. В. Смирнов
斯奈尔	J. L. Snell	Дж. Снелл
斯皮策	F. Spitzer	Ф. Спицер
斯坦因豪斯	H. D. Steihaus	Г. Д. Штейнгаус
斯特林	J. Stirling	Дж. Стирлинг
斯特扎克	D. R. Stirzaker	Д. Стирэакер
斯梯格列尔	M. S. Stigler	С. Стиглер
斯通	M. H. Stone	М. Г. Стоун
斯托特	W. F. Stout	В. Ф. Стоут
所罗门诺夫	R. Solomonov	Р. Соломонов

T

塔尔塔利亚	N. Tartalya	Н. Тарталья
泰切尔	H. Teicher	Г. Тейчер
汤斯凯	M. Donsker	М. Донскер
特普利茨	O. Toeplitz	О. Тёплиц
图尔恰	I. T ulcea	И. Тулча
托德汉特	I. Todhanter	Э. Тодхантер

托德亨特	I. Todhunter	И. Тодхатер
托洛佐娃	T. B. Tolozowa	Т. Б. Толозова

W

威曼	A. Wiman	А. Уиман
威伊达	I. Vajda	И. Вайда
韦布尔	W. Weibull	В. Вейбулл
维尔	J. Ville	Ж. Вилль
维尔斯特拉斯	K. T. W. Weirstrass	К. Т. В. Вейерштрасс
维纳	N. Weiner	Н. Винер
温策尔	A. D. Wentzel	А. Д. Вентцель
温特纳	A. Wintner	А. Винтнер
瓦尔德	A. Wald	А. Вальд
沃尔德	H. Wold	Г. Вольд
沃尔夫	R. Wolf	Р. Вольф
乌沙科夫	W. G. Ushakoff	В. Г. Ушаков
伍拉姆	Woollam	Улам

X

西拉日季诺夫	S. H. Sealaijinov	С. Х. Сираждинов
西格蒙德	D. Sigmund	Д. Сигмунд
希尔伯特	D. Hilbert	Д. Гильберт
希奈	Y. G. Sinaî	Я. Г. Синай
谢瓦斯契亚诺夫	B. A. Sevastyanov	Б. А. Севастьянов
辛钦	A. J. Khintchine	А. Я. Хичин
休伊特	E. Hewitt	Э. Хьюитт

Y

雅可比	G. G. J. Jaccbi	К. Г. Я. Якоби
亚当斯	J. G. Adams	Д. К. Адамс
亚格洛姆	A. M. Jaglom	А. М. Яглом
亚格洛姆	E. M. Jaglom	И. М. Яглом
亚历山大罗夫	A. S. Alexaderdrov	П. С. Алексадров
亚历山大罗娃	N. W. Alexaderdrova	Н. В. Алексадрова
延森	I. L. Iensen	И. Л. Иенсен
伊金格	Ezenger	Изинг

伊藤清	Î to Kiyosi	К. Ито
伊西哈尔		А. Исихар
易卜拉给莫夫	E. A. Ibragemov	И. А. Ибрагимов
尤什克维奇	A. P. Yushikaivici	А. П. Юшкевич

Z

扎克德	J. Jacod	Ж. Жакод
扎克斯	S. Zacks	Ш. Закс
钟开莱	Kai Lai Chung	Чжун Кай -лай
祖布科夫	A. M. Zubkoff	А. М. Зубков
佐洛塔廖夫	W. M. Zolotaileoff	В. М. Золотарёв



记号索引





常用数学符号

$\mathbb{R}=(-\infty,\infty)$ —— 实数的集合, 实直线, 一维欧几里得空间

$\mathbb{R}_+=[0,\infty)$

$\overline{\mathbb{R}}=[-\infty,\infty]$ —— 扩充实直线: $\overline{\mathbb{R}}=\mathbb{R}\bigcup\{-\infty\}\bigcup\{\infty\}$

$\overline{\mathbb{R}}_+=[0,\infty]$

$\mathbb{Q}$ —— 有理数的集合

$\mathbb{Q}_+=\mathbb{Q}\bigcap\mathbb{R}_+$

$\mathbb{R}^{\mathrm{d}}$ ——d 维欧几里得空间

$\mathbb{N}$ —— 自然数: $\{0,1,2,\cdots\}$ 或 $\{1,2,\cdots\}$

$\mathbb{Z}$ —— 整数的集合: $\{0,\pm1,\pm2,\cdots\}$

$\mathbb{C}$ —— 复数的集合

$$(a,b)=\{x\in\overline{\mathbb{R}}:a<x<b\},[a,b]=\{x\in\overline{\mathbb{R}}:a\leqslant x\leqslant b\}$$

$$(a,b]=\{x\in\overline{\mathbb{R}}:a<x\leqslant b\},[a,b)=\{x\in\overline{\mathbb{R}}:a\leqslant x<b\}$$

$\inf X$ —— 集合 $X\subseteq\overline{\mathbb{R}}$ 的下界

$\sup X$ —— 集合 $X\subseteq\overline{\mathbb{R}}$ 的上界

$\inf\limits_{n\geqslant m}x_n$ —— 集合 $X=\{x_m,x_{m+1},\cdots\}$ 的下界

$\sup\limits_{n\geqslant m}x_n$ —— 集合 $X=\{x_m,x_{m+1},\cdots\}$ 的上界

如果$x_n\in\overline{\mathbb{R}},n\geqslant1$, 则

$$\liminf_{n\to\infty}x_n=\underline{\lim_n}x_n\equiv\sup_{m\geqslant1}\inf_{n\geqslant m}x_n,\quad\limsup_{n\to\infty}x_n=\overline{\lim_n}x_n\equiv\inf_{m\geqslant1}\sup_{n\geqslant m}x_n,$$

$$\lim x_n=x\Leftrightarrow\underline{\lim}x_n=\overline{\lim}x_n=x\Leftrightarrow\underline{\lim}x_n\geqslant x\geqslant\overline{\lim}x_n.$$

对于实数

$x^+ = \max(x,0), x^- = -\min(x,0)$

$x^{\oplus} = \begin{cases} x^{-1}, & 若\ x \neq 0, \\ 0, & 若\ x = 0 \end{cases}$

$x \vee y = \max(x,y), x \wedge y = \min(x,y)$

$[x]$ 或 $\lfloor x \rfloor$ —— 不大于 x 的最大整数

$\lceil x \rceil$ —— 大于或等于 x 的最小整数

sign x —— 实数 x 的符号:

$$\text{sign}\ x = \begin{cases} 1, & 若\ x > 0, \\ 0, & 若\ x = 0, \\ -1, & 若\ x < 0 \end{cases}$$

(有时, 当 $x \geqslant 0$ 时, 设 sign $= 1$; 当 $x < 0$ 时, 设 sign $= -1$)

$x_n \to x$, 其中 $n \in \{1, 2, \cdots\}$, 表示 $\lim\limits_n x_n = x$

$x_n \uparrow$ 表示 $x_1 \leqslant x_2 \leqslant \cdots$; $x_n \uparrow x$ 表示 $x_n \uparrow$ 且 $\lim\limits_n x_n = x$

$x_n \downarrow$ 表示 $x_1 \geqslant x_2 \geqslant \cdots$; $x_n \downarrow x$ 表示 $x_n \downarrow$ 且 $\lim\limits_n x_n = x$

对于复数 $z = a + ib$, 其中 $a, b \in \mathbb{R}$, 而 $i = \sqrt{-1}$ 是虚单位

$\bar{z} = a - ib$ —— z 的共轭复数

$|z|$ —— z 的模 $(= \sqrt{a^2 + b^2})$

Re z 和 Im z —— z 的实部和虚部: Re $z = a$, Im $z = b$

对于 $d-$ 维欧几里得空间 $\mathbb{R}^d$

$|x|$ —— $x = (x_1, \cdots, x_d)$ 的欧几里得范数, 即 $\sqrt{x_1^2 + \cdots + x_d^2}$

$x \cdot y$ 或 (x, y) —— $x = (x_1, \cdots, x_d)$ 和 $y = (y_1, \cdots, y_d)$ 的数量积, 即 $x_1 y_1 + \cdots + x_d y_d$

集合论

$A_n \uparrow$ 表示 $A_1 \subseteq A_2 \subseteq \cdots$; $A_n \uparrow A$ 表示 $A_n \uparrow$ 且 $\bigcup A_n = A$

$A_n \downarrow$ 表示 $A_1 \supseteq A_2 \supseteq \cdots$; $A_n \downarrow A$ 表示 $A_n \downarrow$ 且 $\bigcap A_n = A$

$\limsup A_n$, 或 $\overline{\lim} A_n$, 或 $\bigcap\limits_{m \geqslant 1} \left(\bigcup\limits_{n \geqslant m} A_n \right)$ 表示 $\{$有无限多个 $A_n\}$ —— 属于无限多个集合 $A_n (n \geqslant 1)$ 的点的集合

$\liminf A_n$, 或 $\underline{\lim} A_n$, 或 $\bigcup\limits_{m \geqslant 1} \left(\bigcap\limits_{n \geqslant m} A_n \right)$ 表示 —— 属于所有 (仅可能有限个 A_n 除外)

集合 $A_n(n \geqslant 1)$ 的点的集合
I_A 或 $I(A)$ —— 集合 A 的示性函数
$\{\cdots\}$ —— 集合

数 学 符 号

$\ll$ —— 绝对连续
$\sim$ —— 等价
$\perp$ —— 垂直
$f = o(g)$ —— $\lim\left(\frac{f}{g}\right) = 0$
$f = O(g)$ —— $\limsup\left|\frac{f}{g}\right| < \infty$
$f \sim g$ —— $\lim\left(\frac{f}{g}\right) = 1$
$f \asymp g$ —— 比值 $\frac{f}{g}$ 自下与 0 分离, 且自上与 ∞ 分离
$f \circ g$ —— f 与 g 的复合
$f * g$ —— f 与 g 的卷积

